NATIONAL QUALITY FORUM

Memo

TO: Consensus Standards Approval Committee (CSAC)

FR: Helen Burstin, Chief Scientific Officer

Marcia Wilson, Senior Vice President, Quality Measurement

RE: Appeal of Measures for the Readmissions 2015-2017 Project

DA: February 7, 2017

ACTION REQUIRED

The CSAC will review the letters of appeal and this memo in consideration of the appeal. The CSAC will determine whether to uphold measure endorsement for the following measures:

- 0330: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization* (CMS)
- 0506: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization* (CMS)
- 1789: *Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)* (CMS)
- 1891: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization* (CMS)
- 2881*: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)* (CMS)

The following documents are appended to this memo:

1. Appendix A: Appeal Letters from Adventist Health System
2. Appendix B: Measure evaluation summary tables
3. Appendix C: Background on the Appealed Measures

BACKGROUND

In accordance with the NQF Consensus Development Process (CDP), the measures recommended by the Readmissions 2015-2017 Standing Committee were released for a 30-day appeals period, which closed on January 11, 2017. The readmission project remains under the existing appeals process. The National Quality Forum (NQF) has received one appeal of its endorsement of the measures listed above from Adventist Health System. Background information for each of the five measures is provided in Appendix C.

SUMMARY OF APPEALS

The appeals focused on the following issues:

- 0330: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization* (CMS)

- 0506: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization* (CMS)
- 1891: *Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization* (CMS)
    - Summary of Appeal: This measure is used by Centers for Medicare and Medicaid Services (CMS) in the Hospital Inpatient Quality Reporting (HIQR) Program and the Hospital Readmission Reduction Program (HRRP). Information from the HIQR program is publicly reported on the Hospital Compare website and the results of measures in the HRRP are used to determine penalties for excess readmissions. The appellants argue that the use of this measure in the ways directly and materially affects their interests.
    - The appeal was made on the grounds that 1) procedural errors were made that were likely to affect the outcome of the original endorsement decision and 2) that new information or evidence has become available that is reasonably likely to have affected the outcome of the original endorsement decision.
        - Procedurally, the appellants raise concerns that the measure did not meet NQF's standards for reliability and that the member vote to not achieve consensus.
        - The appellants note two new piece of information available. First is a December 2016 report published by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) titled "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." The second item is a New England Journal of Medicine (NEJM) article titled "Should Medicare Value-Based Purchasing Take Social Risk into Account?" published on December 28, 2016.
- 1789: *Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)* (CMS)
    - Summary of Appeal: This measure is used by Centers for Medicare and Medicaid Services (CMS) in the Hospital Inpatient Quality Reporting (HIQR) Program. Information from the HIQR program is publicly reported on the Hospital Compare website. The appellants argue that the use of this measure in the ways directly and materially affects their interests.
    - The appeal was made on the grounds the same grounds as above.
- 2881*: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)* (CMS)
    - Summary of Appeal: This measure will be used by the Centers for Medicare and Medicaid Services (CMS) in the HIQR program. The appellants raise concerns that the public reporting of this measure will directly and materially affect their interests.

- The appeal was made on the same grounds as above.

MEASURE DEVELOPER RESPONSE TO THE APPEAL

We thank NQF for the opportunity to respond to the recent appeal by Adventist Health System regarding the endorsement of measures #0330, #0506, #1789, #1891, and #2881 in the All-Cause Admissions and Readmissions project:

- NQF #0330: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Heart Failure (HF) Hospitalization
- NQF #0506: Hospital 30-day, All-cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
- NQF #1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
- NQF #1891: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
- NQF #2881: Excess Days in Acute Care (EDAC) After Hospitalization for Acute Myocardial Infarction (AMI)

The appeal is based on the claim of two procedural errors and the availability of new information or evidence. We address only two of the issues raised in this response below.

I. Measure Reliability

The appellant asserts that "we believe that the recommendation and subsequent endorsement of several of these measures was inconsistent with NQF's Scientific Acceptability criterion for reliability."

To support the appeal, several sources regarding reliability and standards for approaches to interpreting statistics measuring reliability are cited. Our position is that the measures under appeal meet this subcriterion, and that no procedural error occurred. We offer three points in a brief rebuttal below and a detailed explanation of our rationale on pages 4-7.

First, in their critique of the results CMS presented for measure score reliability, the appellant cites NQF's guidance on interpretation of **data element reliability**. Because these measures are calculated from claims submitted by hospitals and other providers, adjudicated by CMS, and stored electronically, the reliability of the data is extremely high. When the measures are computed on the same set of admissions, for the same providers, using the same time period, precisely the same results are obtained. That is, these are deterministic measures, reproducible

by any third party, and thus demonstrably meet the standard described by NQF under item 2a2. We maintain that the NQF's measure submission forms *offer no guidance* on the interpretation test of **measure score reliability**, including the test used by CMS, the intraclass correlation coefficient, ICC[2,1]. This is a test-re-test method and NQF's guidance on interpretation of data element reliability does not apply.

Second, the appellant cites several sources that use signal-to-noise ratios to evaluate provider measures. The appellant suggests that these are suitable approaches to assessing measures score reliability, and that the critiqued measures don't meet the standards for signal-to-noise reliability. We maintain that signal-to-noise ratio is useful for some purposes, but **signal-to-noise ratio is a provider level metric**, which assesses reliability separately for each provider's measure score. This metric is then typically averaged across all providers to create a measure reliability score. This is not consistent with standard approaches to evaluating measure reliabilities. Moreover, because signal-to-noise is the ratio of between unit variation (signal) to total between unit plus within unit variation (precision), a measure can be very imprecise at the unit level and still have a high signal-to-noise ratio, if there is large between unit variation. Conversely, a measure can be extremely precise for each unit, but have very low signal-to-noise reliability, if there is no between unit variation. For this reason, **signal-to-noise ratio is not consistent with the reliability metric we report, ICC[2,1]**. In the details at the end of this memo we report on a simulated dataset with high signal-to-noise reliability and low ICC[2,1]. Thus, we maintain that the standards that are referenced for signal-to-noise ratio do not apply to ICC[2,1].

Third, since, as noted above, there is no NQF guidance for standards of test-retest reliability, and the standards cited for signal-to-noise ratio do not apply, **other guidelines or reference values for ICC[2,1] should be used**. In the absence of empirically supported standards, our position is that 'acceptability' depends on context. For simple concepts or constructs, such as a patient's weight, the expectation is that the test-retest reliability of a measure of that construct should be quite high. However, for complex constructs, such as clinical severity, patient comorbidity, or symptom profiles used to identify a condition or clinical state, reliability of measures used to define these constructs is quite a bit lower. In this memo we offer several examples of the reliability of measures of complex constructs using the ICC[2,1]. These examples provide the necessary context for interpreting the acceptability of ICC[2,1] values in the ranges found for the readmission measures. **These empirical findings indicate that our reported ICC[2,1] values are consistent with those in similar contexts.**

II. New Publications Related to the Use of SES in Measure Risk-Adjustment Models

We have reviewed the two recent studies mentioned in the appeals' letter. Both the ASPE

report[1] and the NEJM article[2] address the importance of social factors in quality measurement and pay-for-performance programs. We have long acknowledged and agree with the conclusion of both studies that socially disadvantaged groups, such as those earning a low-income, members of some racial or ethnic minority groups or those living with a disability, are at greater risk of poor health and health outcomes. However, we disagree that either study provided new evidence that was meaningfully different than the evidence available to the committee during their deliberations on these measures.

The ASPE report and the NEJM article restate the NQFs recommendation that measures should be examined *individually* to determine if adjustment for social factors is appropriate (ASPE 2016: 15; NEJM 2017: 2). When determining whether adjustment is warranted, developers were instructed to consider the conceptual relationship between the SDS factor and the outcome as well as the empirical relationship. Although we found that both observed and adjusted readmission rates are higher on average for hospitals serving a large proportion of patients who were dual eligible and those living in a census block group with low AHRQ SES Index, we have also shown that many hospitals serving a high share of socially disadvantaged patients achieve high performance scores on the readmission measures (see, e.g., Bernheim et al. 2016. "Accounting for Patients' Socioeconomic Status Does Not Change Hospital Readmission Rates." Health Affairs 35(8): 1461-1470). The authors of the NEJM article also found that risk adjusting for indicators of SES or of race does not explain away performance differences between hospitals serving low- and high-proportions of beneficiaries with these indicators, which aligns with our findings presented to the Committee. Neither publication offered new relevant information that was not available to the Committee during their deliberations.

The ASPE report does not recommend risk-adjustment of readmission measures with SES risk variables. However, the report does recommend consideration of stratifying hospitals into peer groups *after* measure calculation for the purpose of payment calculation *rather than* adjusting measures at the patient level: "Hospitals would be judged only against their peers, and penalties would be assessed based on the average performance within each group rather than the average performance overall" (ASPE report 2016: 82). The recent 21st century CURES laws align with this recommendation and direct CMS to stratify hospitals for the purpose of determining the payment adjustment factor within the Hospital Readmission program (HRRP). This represents a change to the use of the measure within a pay-for-performance program but not a change to the measure itself.

We agree with the appellants comment that patient-level stratification of readmission rates (in contrast to stratifying hospitals into peer groups) could serve to illuminate disparities within hospitals of quality of care for beneficiaries with social risk factors. However, the NQFs guidance

[1] Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE), "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." December 2016, https:/ /aspe.hhs.gov/ pdf-report/ report-congress-social-risk-factors-and-performance-under-medicares-value-based-purchasing-programs

[2] Joynt, Karen E., De Lew, Nancy, Sheingold, Steven H., Conway, Patrick H., Goodrich, Kate, and Epstein, Arnold M. (2017) "Should Medicare Value-Based Purchasing Take Social Risk into Account?" *New England Journal of Medicine.* http://www.nejm.org/doi/full/10.1056/NEJMp1616278#t=article

to measure developers for the SDS trial period was to present stratified results to the committee only for measures that included SES indicators in the measure risk model. Therefore, we did not submit stratified measure results.

We agree with the appellant's comment, the ASPE report, and the NEJM article recommendations to measure and monitor quality of care for vulnerable populations, but adding patient-level risk-adjustment to the readmission measures is not a means to do so. The rationale behind the development of equity measures is to illuminate disparities and create incentives to reduce them, improve care for vulnerable populations, and promote greater transparency for consumer choice. An example of such an initiative is the graphical tool "Mapping Medicare Disparities" provided by the Office of Minority Health (OMH), which identifies geographical areas of disparities between subgroups of Medicare beneficiaries (e.g., dual eligible vs. non-dual eligible beneficiaries) on its webpage (https://www.cms.gov/about-cms/agency-information/omh/OMH-Mapping-Medicare-Disparities.html). CMS supports these and other initiatives to highlight disparities and promote greater equity in health care delivery and patient outcomes. CMS remains committed to developing alternative ways to measure and report disparities and to promote equity in care and outcomes among beneficiaries.

Additional Details on Measure Reliability

The appellant cites the NQF subcriterion 2a2, which states the criterion for reliability:

> *"Reliability testing demonstrates that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise." (emphasis added).*

Notably, **this subcriterion has a footnote:**

> *Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).*

Many concerns about reliability of measures and measure attributes arise because of the multiple definitions of reliability and the multiple standards available in the literature. In this footnote to 2a2 we see a long list of somewhat exclusive types of reliability listed. Here we discuss three metrics of reliability that are relevant.

COMPUTED SCORE RELIABILITY

The appellant claims that the measures reported do not meet the standards of Subcriterion 2a2, which specifically requires that "*measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period*", a standard which according to the footnote applies to *the data elements and the*

computed measure score. We will refer to this as the computed measure score reliability. This reliability can be low for measures that rely for instance on surveys (where respondents can be inconsistent with responses), data abstraction (which can introduce errors) or collecting new clinical data (which has measurement error), but is typically high for measures that rely on existing claims data. For the measures appealed, all data used to calculate the measures are derived from adjudicated and finalized Medicare claims, which are submitted and stored electronically. The reliability of such data is extremely and uniformly high. And, given the dataset of data elements of demonstrated reliability, when the measures are computed on the same set of admissions, for the same providers, using the same time period, precisely the same results are obtained. That is, these are deterministic measures, reproducible by any third party, and demonstrably meet the standard of 2a2; they produce exactly the same results nearly 100% of the time.

SIGNAL-TO-NOISE RELIABILITY

The appellant memo cites several sources that use signal-to-noise ratios to evaluate provider measures. Signal-to-noise is a type of reliability, but it is distinct from both computed score and test-retest reliability. *This notion of reliability is not related to test-retest reliability*. It is listed in the NQF Subsection 2a2 footnote above as an example of precision, but strictly speaking, it is not a measure of precision. Rather, measures of signal-to-noise (there are several) reflect the ratio of between unit variation (signal) to total variation (between unit plus within unit, where the within unit variation reflects precision). As noted earlier, a measure can be very imprecise at the unit level and still have a high signal to noise ratio, if there is large between unit variation; conversely, a measure can be extremely precise at the unit level but have low signal-to-noise ratio, if there is no between unit variation. Moreover, it is unit level metric, calculated separately for each provider, typically averaged to create a 'measure' reliability. For both of these reasons, we do not think it is an appropriate measure of *measure* reliability; instead we use test-retest reliability.

Because the signal-to-noise ratio measure does not equate to the test-retest reliability measure, the same conventional thresholds do not apply to both. Thompson et al, cited by the appellant, uses 0.7 as a threshold, and justifies this with references to other authors who used it; there seems to be no empiric justification, as we provide below for test-retest reliability. To demonstrate the distinction between this approach and test-retest, we simulated a dataset (available) to demonstrate the difference between ICC[2,1] and Signal-to-Noise ratio. In this simulated dataset, which includes 100 hospitals with mean rate of 20%, and an average of 50 patients per hospital we found ICC[2,1] = 0.20 and the average Signal/Noise ratio = 0.73. This example demonstrates explicitly that the signal/noise ratio is distinct from ICC[2,1], and that the papers, standards and reports referenced by the appellant do not apply.

TEST-RETEST RELIABILITY

The measure of test-retest reliability used to assess the measures is a specific statistic known as ICC[2,1], which is analogous to the more familiar □ but appropriate for continuous measures. It compares two repeated measures on each provider for agreement; it is a conservative measure

of test-retest reliability, because it assumes that the multiple measurements are drawn from a larger sample of tests, and that the measured providers are drawn from a larger sample of providers. This reliability does not refer to the reliability of the data elements or the precision of the estimates, the two criteria mentioned in 2a2, but rather the reliability of the risk-adjusted measure score. Note that ICC[2,1] is also distinct from the conventional "intra-class correlation", which the ratio of between unit variation to the total variation.

The appellant then references the ICC[2,1] values reported for the challenged measures. Note that these are reported as additional reliability testing, per the footnote to 2a2. No standards are given for the types of reliability listed in the footnote. In particular, ICC[2,1] evaluates the reliability of the measure with respect to different data samples (split samples which include data from separate groups of patient admissions). Moreover, guidelines for the specific ICC[2,1] statistic are of limited availability. The appellant cites only a single source for evaluating ICC[2,1], Rousson et al, who however simply cite Lee et al; Lee et al in turn reference Burdock et al without comment. However, Burdock et al mention 0.75 without any justification, and for a different statistic:

> *"$R=ut^2/(ue^2 + ut^2)$ is based on the assumptions that the observers are fixed and that there is no interaction between observer and subject. Apart from considerations of the other components of the model, a minimum requirement of the instrument is that R be large, meaning that is should be as close to unity as possible. A high intraclass correlation coefficient, e.g. $R \geq .75$".*

Note that Burdock et al provides no empiric justification, and moreover, are discussing a reliability metric that is not ICC[2,1], but something more similar to a signal-to-noise ratio.

If there is no evidence to support the 0.75 value, the question remains how to best determine what is an 'acceptable' level of inter-rater reliability. Some may still use the Landis & Koch (Landis, Koch 1977) convention to argue that CMS hospital measures have poor reliability, that something in the 0.61-0.80 range might be more appropriate ("substantial"), which coincides with a common instinct to think of 60% as "passing" and 80% as "above average." However, conventions are by definition flexible; to quote Landis & Koch, which NQF has mentioned as a guideline:

> *In order to maintain consistent nomenclature when describing the relative strength of agreement associated with kappa statistics, the following labels will be assigned to the corresponding ranges of kappa …* ***Although these divisions are clearly arbitrary****, they do provide useful "benchmarks" for the discussion of the specific example in Table 1*

Thus, even these guidelines, which have been widely adopted, were originally stated as arbitrary. Their usefulness has derived largely from their consistency with findings in a very large range of research fields over the four decades since their original publication. However, this does not make them final standards of 'acceptability'.

Therefore, our position is that 'acceptability' depends on context. For example, if we were

measuring adolescent weight twice with the same scale, and assessing whether the weights were above a certain threshold, we would expect the two measurements to agree almost exactly (ICC[2,1] ~ 1); otherwise, we would discard the scale. At the other extreme, if we were measuring a latent personality trait such as a personality disorder, we would expect a much lower level of agreement. In fact, Nestadt et al assessed ICCs for several standard tools for assessing personality disorder and found test-retest reliabilities in the range of 0.06-0.27 (Nestadt 2012). (Notably, Nestadt et al conclude that these tools "may still be useful for identifying [personality disorder] constructs.")

Thus, we would argue that **one should adopt for 'acceptable' level of ICC[2,1] a standard that is consistent with that in known, familiar, and related contexts**. The current context is measuring provider quality, or specifically provider propensity to provide appropriate care as measured by subsequent outcomes. We identified several studies, which we think support the Landis & Koch guidelines when assessing test-retest reliability in the context of hospital measurement.

- Hall et al calculated test-retest reliability for determining comorbidities from chart abstraction [Hall et al]. In this study, multiple abstracters abstracted the same charts and the results were used to calculate four different common comorbidity scores. For three of the indices, test-retest reliabilities ranged from 0.59-0.68, with the fourth (the Charlson comorbidity score) achieving 0.80. We would argue that chart abstraction, with test-retest reliabilities in the 'moderate' to 'substantial' range, should be inherently more reliable than measuring hospital quality.
- Cruz et al report reliabilities for collecting risk factor information from patients presenting to an emergency department with potential acute coronary syndrome (ACS) [Cruz et al]. Each patient was queried twice, once by a clinician and once by a trained research assistant, and the reliabilities for a range of risk factors were calculated; these ranged from 0.28 (associated symptoms) to 0.69 (cardiac risk factors), with all other factors in the 0.30-0.56 range.
- Hand et al report test-retest reliabilities for bedside clinical assessment of suspected stroke [Hand et al]. Pairs of observers independently assessed suspected stroke patients; findings were recorded on a standard form to promote consistency. The reliabilities were calculated for the full range of diagnostic factors: for vascular factors reliabilities ranged from 0.47-0.69 with only four of eight above 0.6; for history they ranged from 0.37-0.65 with only five of 12 above 0.6; other categories were similar (though reliability=1 for whether the patients were conscious).

These contexts are intuitively similar to that of measuring hospital quality, and moreover suggest that the guidelines of Landis & Koch are appropriate for areas of clinical care.

SUMMARY

The appealed measures do meet the standard of high computed score reliability specified in NQF guideline section 2a2. Signal-to-noise reliability, while useful, is not a metric of scale reliability, and is distinct from test-retest reliability, and any conventional thresholds do not

necessarily apply to ICC[2,1]. Accepted standards for ICC[2,1] are not available, but an examination of test-retest reliability in contexts that are intuitively similar to that of provider quality measurement finds values that are consistent with both the alternative guidelines and with CMS measures. Reliability testing in hospital quality measurement should be interpreted in context, and the evidence we present refutes that 0.7 is a minimal acceptable reliability value for test-retest reliability of complex clinical constructs such as symptomatology, health risk factors, comorbidity, or hospital performance on patient outcomes.

STANDING COMMITTEE AND NQF RESPONSE TO THE APPEAL

Standing Committee Response

The Standing Committee did not have an additional response to the concerns raised by the appellants. During the endorsement process, the Standing Committee deliberated extensively on the potential need to adjust these measures for social risk factors. In response to public comment raising concerns that these measures do not include social risk factors, the Committee reiterated that their focus was on evaluating the measure specifications and testing submitted by the measure developer.

- The Committee recognized that there continues to be limitations in the available data elements to capture unmeasured clinical and socio-demographic risk.
- Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between SDS factors and readmissions.
- The Committee recommended the endorsement of the measures without adjustment for social risk factors based on the small effect size of those factors in the analyses put forth by the developer. The Committee stressed the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards.
- The Committee recognized the need to ensure facilities serving vulnerable populations are not penalized unfairly while, at the same time, balancing concerns about worsening healthcare disparities.
- The Committee recommends a reassessment of the availability of SDS variables and a reexamination of these measures through the NQF annual update process.

NQF Response

Standards for Reliability:

NQF does not maintain a set standard for reliability. When developers use test-retest reliability to assess the Intra-class Correlation Coefficient (ICC), NQF provides information on the conventions put forth by Landis and Koch in the preliminary analysis developed for each measure. However, the Standing Committee retains the ability to make their own assessments on the reliability of a measure.

Results of Member Vote:

Once a project standing committee has reviewed all of the comments submitted during the public and member commenting period and made any revisions to the draft report, members of NQF vote on the candidate standards that are recommended by the committee. All candidate consensus standards that are recommended with the results of the voting by the membership will proceed to the next step in the consensus development process: decision by the Consensus Standards Approval Committee (CSAC). NQF staff provide a summary of Member voting to the CSAC. If the member voting does not reach consensus >60%, CSAC has the option to request a re-vote or an all-member meeting.

The memo to CSAC on the Readmissions 2015-2017 project highlighted the member voting results. The memo noted that one of the recommended measures was approved with 67% or higher. The memo also stated that Representatives of 19 member organizations voted; no votes were received from Consumer, Supplier/Industry, or Public/Community Health Agency Councils. Detailed breakdowns of the vote on each memo were provided in an appendix.

The CSAC did not request a re-vote or an all-member meeting on the voting results of #0330, #0506, #1789, #1891, or #2881.

APPENDIX A: APPEAL LETTER

Measure 2881 Appeal Request

Adventist Health System, Submitted January 11, 2017

Adventist Health System (AHS) wishes to appeal the decision to endorse the excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI) (NQF# 2881).
We believe our interests will be directly and materially affected by this recently endorsed consensus standard because will be used by the Centers for Medicare and Medicaid Services (CMS) in the Hospital Inpatient Quality Reporting (IQR) Program. This program has a substantial impact on AHS facilities. HIQR measure results are publicly reported and affect public perception of AHS hospital facilities.

We wish to appeal the endorsement of this measure on grounds that 1) procedural errors were made that were likely to affect the outcome of the original endorsement decision and 2) on the grounds that new information or evidence has become available that is reasonably likely to have affected the outcome of the original endorsement decision.
It is the view of AHS that two significant procedural errors were made in the decision to endorse this measure.

First, the Standing Committee should not have found that this measure meets the NQF's standard for reliability. The developer used a "test-retest" approached to assess reliability. The agreement between two RSRRs, as measured by Intra-class Correlation Coefficient (ICC), was 0.54. The measure developer, in its response to comments, cited a convention that "describes the ICC values as moderate (0.41-0.60) for this measure" (Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977; 33:159-174). AHS agrees with Landis and Koch [1977] that "[a]though these divisions are clearly arbitrary, they do provide useful 'benchmarks' for the discussion of [a] specific example […]" Furthermore, we agree with the developer that the ICC values of this measure could be described as "moderate" under the "benchmarks" put forward by Landis and Koch [1977]. However, AHS believes that NQF committees should only assess a measure as meeting NQF standards for reliability if that measures meets a threshold of reliability commensurate with the impact of its current or prospective use. It is our opinion that achieving a "moderate" benchmark of reliability is not sufficient for the endorsement of substantially impactful measures. We find measures that are used in public reporting or payment programs, such as the HIQR and HRRP, to be substantially impactful. Hence, AHS believes that for such measures to be awarded endorsement they should first be assessed as meeting a reliability benchmark or "strength of agreement" that is "substantial" according to Landis and Koch [1977]. Thus, we conclude that, according to the Landis and Koch [1977] convention cited by the developer, the "substantial" reliability "benchmark" for this measure would be an ICC value of 0.61-0.80. In other words, AHS believes that, by the developer's own scale, this measure should have achieved an ICC of at least 0.61 to meet the NQF's standard for reliability.

Second, this vote did not achieve consensus among the NQF member organizations that cast votes during the endorsement proceedings. Six members voted in favor of endorsement of the measure and seven members voted against endorsement of the measure. That is an approval rate of 46 percent. AHS believes that a member voting approval rate of 46 percent is insufficient for NQF endorsement. We think it is also worth pointing out that only three out of the eight measure councils had more than two members cast votes. Of these three councils, only one approved of the measure. We find it alarming that a measure can achieve NQF endorsement

despite receiving more votes of disapproval than approval. It is our opinion that the NQF's status as the "gold standard" of quality measurement and as a consensus standard body (as defined by the Office of Management and Budget) could be in serious jeopardy if this trend persists.

It is also the view of AHS that two pieces of new information have become available since the CSAC made its endorsement decision that are reasonably likely to affect the outcome of the original endorsement decision.

The first item was a December 2016 report published by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) titled "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." The report concluded that "social factors are powerful determinants of health. In Medicare, beneficiaries with social risk factors have worse outcomes on many quality measures, including measures of processes of care, intermediate outcomes, outcomes, safety, and patient/consumer experience, as well as higher costs and resource use. Beneficiaries with social risk factors may have poorer outcomes due to higher levels of medical risk, worse living environments, greater challenges in adherence and lifestyle, and/or bias or discrimination. Providers serving these beneficiaries may have poorer performance due to fewer resources, more challenging clinical workloads, lower levels of community support, or worse quality."

In addition, the report recommended that "measuring and reporting quality for beneficiaries with social risk factors, setting high, fair quality standards for all beneficiaries."

The second item was a New England Journal of Medicine (NEJM) article titled "Should Medicare Value-Based Purchasing Take Social Risk into Account?" that was published on December 28, 2016. This article noted that "beneficiaries with social risk factors had worse outcomes on many quality measures, regardless of the providers they saw, and dual enrollment status was the most powerful predictor of poor outcomes." In addition, the article highlighted that "providers that disproportionately served beneficiaries with social risk factors tended to have worse performance on quality measures." The article also recommended that "we should measure and report quality of care for beneficiaries with social risk factors."

AHS believes that the HHS ASPE report and NEJM article highlight what the NQF's Readmission Committee stressed as "the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards." It is our view that these reports represent advancements in the field that the committee suggested would necessitate reevaluation. Therefore, endorsement of this measure should be revoked because the information presented by these reports is reasonably likely to have affected the original endorsement decision.

Measure 0330, 0506, 1789, 1891, 2881 Appeal Request
Adventist Health System, Submitted January 20, 2017

To Whom It May Concern:

I am writing on behalf of Adventist Health System (AHS) to appeal the decision to endorse the following NQF Readmission measures:

- NQF #0330: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR)

Following Heart Failure (HF) Hospitalization
- NQF #0506: Hospital 30-day, All-cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
- NQF #1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
- NQF #1891: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
- NQF #2881: Excess Days in Acute Care (EDAC) After Hospitalization for Acute Myocardial Infarction (AMI)

We believe our interests are directly and materially affected by these recently endorsed consensus standards because they are used or are proposed to be used by the Centers for Medicare and Medicaid Services (CMS) in the Hospital Inpatient Quality Reporting (HIQR) Program and the Hospital Readmission Reduction Program (HRRP). These federal quality measurement programs are substantially impactful. HIQR measure results are publicly reported and thereby affect public perception of AHS hospital facilities. HRRP measures results are used to adjust payments that AHS hospital facilities receive from Medicare.

A recent study, titled "Reliability of 30-Day Readmission Measures Used in the Hospital Readmission Reduction Program," that was published in the Health Services Research journal, concluded that "[m]any of the RSRRs employed by the HRRP are unreliable" and "few hospitals have acceptable reliability on all measures for which they are assessed by HRRP." Furthermore, Adventist Health System — NQF Readmission Measures Endorsement Appeal the study found that "one quarter of payments [penalties] for excess readmissions are associated with unreliable RSRRs."

According to the authors, for many hospitals "[HRRP] penalties are likely the result of statistical noise and unlikely to provide constructive information about areas needing improvement." AHS believes that one quarter of the payment penalties tied to readmissions measures is substantial and material.

We wish to appeal the endorsement of these measures on the grounds that procedural errors were made that were likely to affect the outcome of the original endorsement decision. We also wish to appeal the endorsement of these measures on the grounds that new information or evidence has become available that is reasonably likely to have affected the outcome of the original endorsement decision.

<u>Procedural Errors</u>
It is the view of AHS that two procedural errors were made in the decision to endorse these measures.

First, we believe that the recommendation and subsequent endorsement of several of these measures was inconsistent with NQF's Scientific Acceptability criterion for reliability.

Second, we believe that the Consensus Standards Approval Committee (CSAC) did not appropriately consider the results of the NQF Member Voting step of the NQF Consensus Development Process (CDP) before moving forward with its recommendation to endorse these measures.

<u>New Information or Evidence</u>
It is also the view of AHS that two pieces of new information have become available since the CSAC made its endorsement decision that are reasonably likely to affect the outcome of the

original endorsement decision.

The first item was a December 2016 report published by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) titled "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs."

The second item was a New England Journal of Medicine (NEJM) article titled "Should Medicare Value-Based Purchasing Take Social Risk into Account?" that was published on December 28, 2016.

Procedural Error — Reliability

Criterion 2 of NQF's Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement specifies that "[m]easures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria."

Subcriterion 2a2 requires that "[r]eliability testing demonstrates that the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise."

A RAND Corporation Technical Report titled "The Reliability of Provider Profiling: A Tutorial," describes reliability as follows:

> Conceptually, it is a ratio of signal to noise. The signal in this case is the proportion of the variability in measured performance that can be explained by real differences in performance. A reliability of zero implies that all the variability in a measure is attributable to measurement error. A reliability of one implies that all the variability is attributable to real differences in performance.

Using simpler terms, a study published in the Annals of Thoracic Surgery notes that "reliability of 0.8 means that 80% of the variance in outcomes is due to true differences in performance while 20% of the variance is attributable to statistical 'noise' or measurement error."

AHS is appealing the endorsement of several readmissions measures recently endorsed by NQF because we believe they were misjudged by the Standing Committee as having met Subcriterion 2a2. In particular, we find that the Committee used a minimum reliability level that is too low.

As specified in the Draft Report for Voting, the reliability of **NQF# 0330: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Heart Failure (HF) Hospitalization** was tested as follows:

> The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. A total of 1,210,454 admissions over a 3-year period were examined, with 604,022 in one sample and 606,432 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples. The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.58.

As specified in the Final Report for Voting, the reliability of **NQF #1891: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization** was tested as follows:

> The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered to be an appropriate method of testing reliability. A total of 925,315 admissions over a 3-year period were examined, with 461,505 in one sample and 463,810 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples. The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.48.

As specified in the Draft Report for Voting, the reliability of **NQF #2881: Excess Days in Acute Care (EDAC) After Hospitalization for Acute Myocardial Infarction (AMI)** was tested as follows:

> The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. For test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, each hospital twice was measured twice, each time using an entirely distinct set of patients. The developer states that the extent to which the calculated measures of these two subsets agree is evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, the developer calculated the intra-class correlation coefficient (ICC) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977). A total of 496,716 admissions were examined, with 248,358 in each sample. The agreement between the two EDAC values for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.54.

In response to AHS' previous comments on measure #0330 the developer noted:

> We used the Inter-Class Correlation (ICC) method to establish the reliability of the measure score. Our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally comparing the agreement between the two resulting performance measures across hospitals (Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.). This is a purposefully conservative approach to assessing reliability and traditional thresholds for acceptability do not apply to interpreting these results. The minimally acceptable threshold noted by AHS is not appropriate for this particular analytic approach. We have cited the more appropriate convention, which describes the ICC values as moderate (0.41-0.60) for this measure (Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977; 33:159-174).

AHS wishes to highlight that Subcriterion 2a2 specifically requires that the results of reliability testing demonstrate that measures can reproduce "the same results a high proportion of the time when assessed in the same population in the same time period." We find that ICC results of 0.58, 0.48, or 0.54 do not demonstrate a level or reliability or repeatability that can be accurately described as producing the "same results a high proportion of them time." For this reason, it is our view that Measures #0330, #1891, and #2881 should not have passed Criterion 2.

According to Rousson et al., in the paper cited by the developer as informing its approach to reliability testing, "a good reliability is attained if the lower bound of the 95 per cent confidence interval is at least 0.75."

Adams, in the previously referenced RAND report, notes that "[p]sychometricians use a rule of thumb of 90 percent for drawing conclusions about individuals [but] lower levels (70-80 percent) are considered acceptable for drawing conclusions about groups."

The National Research Council's Committee on Performance of Military Personnel has reported that for personnel performance measures "[a]ccepted standards in the field are vague and depend on the characteristic being measured: generally speaking, reliabilities of .6 to .7 are considered marginal, .7 to .8 acceptable, .8 to .9, very good, and above .9 excellent."

According to Thompson et al., 0.70 is "a commonly used benchmark for acceptable reliability, […] for group-level comparisons"

Shih and Dimick note that "[a] commonly used cutoff for acceptable reliability when comparing performance of groups is 0.7."

Furthermore, "the more appropriate convention" cited by the developer was described, in the same paper, by Landis and Koch as "clearly arbitrary." Even taken at face value, the Landis and Koch benchmarks describe reliability kappas of 0.41-0.60 as "Moderate" in terms of "Strength of Agreement." AHS believes that "Moderate" reliability does not align with NQF's criteria.

We think it is clear that the reliability testing results for measures #0330, #1891, and #2881 do not demonstrate that the measures scores are "repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period."

Measure #0330's tested ICC score of 0.58 suggests that only 58 percent of the variation in hospital performance is due to true differences in quality (signal) while 42 percent of the variation is due to measure error (noise).

Measure #1891's tested ICC score of 0.48 suggests that only 48 percent of the variation in hospital performance is due to true differences in quality (signal) while 52 percent of the variation is due to measure error (noise). **AHS wishes to highlight that this reliability score would seem to indicate that this measure does not produce the same results a majority of the time, let alone a *high proportion* of time.**

Measure #2881's tested ICC score of 0.54 suggests that only 54 percent of the variation in hospital performance is due to true differences in quality (signal) while 48 percent of the variation is due to measure error (noise).

We believe that there may be some confusion about reliability due to a lack of guidance from NQF as to what testing results specifically demonstrate sufficient reliability. AHS believes that the Patient Safety Standing Committee may have been highlighting a similar issue when it referenced, in its most recent report, concerns "about insufficient guidance on how to assess measure reliability and validity."

Procedural Error — Voting

It is the view of AHS that the following measures did not achieve consensus during the NQF Member Voting step of the NQF Consensus Development Process (CDP). According to the memo that asked the Executive Committee to ratify the CSAC's recommendation to endorse all 16 measures of the All-Cause Admissions and Readmissions Project 2015-2017, only "[o]ne of the recommended measures was approved, with 67 percent approval or higher by the councils." There was no discussion about why the CSAC chose to recommend all 16 measures despite the fact that only one of the measures achieved greater than 67 percent approval of NQF members. Highlighted below are five substantially impactful measures that did not achieve a simple majority approval rate among NQF members.

- NQF# 0330: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Heart Failure (HF) Hospitalization
    - Approval Rate = 45 Percent
- NQF #0506: Hospital 30-day, All-cause, Risk-Standardized Readmission Rate (RSRR) Following Pneumonia Hospitalization
    - Approval Rate = 50 Percent
- NQF #1789: Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
    - Approval Rate = 50 Percent
- NQF #1891: Hospital 30-day, All-Cause, Risk-Standardized Readmission Rate (RSRR) Following Chronic Obstructive Pulmonary Disease (COPD) Hospitalization
    - Approval Rate = 45 Percent
- NQF #2881: Excess Days in Acute Care (EDAC) After Hospitalization for Acute Myocardial Infarction (AMI)
    - Approval Rate = 46 Percent

AHS believes that a member voting approval rate of 60 percent or less is insufficient for NQF endorsement. We think it is also worth pointing out that for all five of the above measures only three out of the eight measure councils had more than two members cast votes. AHS questions how this can be acceptable for endorsement. We find it alarming that a measure can receive NQF endorsement despite not achieving a majority approval rate among NQF members. It is our opinion that the NQF's status as the "gold standard" of quality measurement and as a consensus standard body, as defined by the Office of Management and Budget, could be in serious jeopardy if this trend persists.

New Information or Evidence — Social Risk Factors

AHS believes that two pieces of new information have become available since the CSAC made its endorsement decision that are reasonably likely to have affected the outcome of the original endorsement decision.

The first item was a December 2016 report published by the U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation (ASPE) titled "Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs." The report included the following findings regarding the Hospital

Readmissions Reduction Program:

> Dually-enrolled beneficiaries had significantly greater odds of readmission than non-dually-enrolled beneficiaries within hospitals, an effect that was relatively similar across hospitals.
>
> There was also a significant hospital effect, suggesting that safety-net hospitals have other unmeasured differences in beneficiary characteristics, provide poorer-quality care to prevent readmissions, or face other barriers that might be related to the availability of resources or community supports.

In addition, the report recommended that:

> readmission rates stratified by social risk should be developed and considered for hospital preview reports and public reporting in places such as Hospital Compare, so that hospitals, health systems, policymakers, and consumers can see and address important disparities in care.

The second item was a New England Journal of Medicine (NEJM) article titled "Should Medicare Value-Based Purchasing Take Social Risk into Account?" that was published on December 28, 2016.

This article noted that:

> beneficiaries with social risk factors had worse outcomes on many quality measures, regardless of the providers they saw, and dual enrollment status was the most powerful predictor of poor outcomes.

In addition, the article highlighted that "providers that disproportionately served beneficiaries with social risk factors tended to have worse performance on quality measures."

The article also recommended that "we should measure and report quality of care for beneficiaries with social risk factors."

AHS believes that the HHS ASPE report and NEJM article highlight what the NQF's Readmission Committee stressed as "the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards."

It is our view that these reports represent advancements in the field that the committee suggested would necessitate reevaluation. Therefore, endorsement of these measures should be withheld until they have demonstrated sufficient risk adjustment and/or stratification for social risk factors.

In conclusion, AHS believes that the endorsement of measures #0330, #0506, #1789, #1789, #1891 and #2881 should be withdrawn due to the procedural errors and new information cited above.

Sincerely, Richard E. Morrison
Adventist Health System
Rich.Morrison@ahss.org
407-357-2377

References

Adams, J. L. 2009. *The Reliability of Provider Profiling: A Tutorial*. Santa Monica, CA: RAND Corporation.

All-Cause Admission and Readmissions 2015-2017: Draft Report for Voting [monograph on the Internet] Washington, D.C.: National Quality Forum; 2016. Oct., [cited 2017 Jan 19]. Available from: Consensus Development Process [Web page on the Internet] Washington, DC: National Quality Forum; c2016 [cited 2017 Jan 19]. Available from:www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx

Joynt, K.E., Lew, N.D., Sheingold, S.H., Conway, P.H., Goodrich, K., and Epstein, A. M. 2016 Dec 28. "Should Medicare Value-Based Purchasing Take Social Risk into Account?" *New England Journal of Medicine* [Epub ahead of print].

Landis, J.R. and Koch, G.C. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33:159-174.

Measure Evaluation Criteria [Web page on the Internet] Washington, DC: National Quality Forum; c2016 [cited 2017 Jan 19]. Available from: http://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx

Patient Safety 2016: Draft Report for Member Vote [monograph on the Internet] Washington, D.C.: National Quality Forum; 2016. Nov., [cited 2017 Jan 19]. Available from: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83819

Shih, T. and Dimick, J. 2014. "Reliability of Readmission Rates as a Hospital Quality Measure in Cardiac Surgery." *Ann Thorac Surg* 97 (4): 1214-1218.

Thompson, M. P. 2016. "Reliability of 30-Day Readmission Measures Used in the Hospital Readmission Reduction Program." *Health Services Research Journal* 51 (6): 2095-2114.

Memo to Consensus Standards Approval Committee Re: All-Cause Admissions and Readmissions 2015-2017 Project [memo on the Internet] Washington, D.C.: National Quality Forum; 2016. Nov., [cited 2017 Jan 19]. Available from: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=83843 Adventist Health System — NQF Readmission Measures Endorsment Appeal Page 10 January 20, 2017

Memo to Executive Committee Re: Ratification of Measures for the All-Cause Admissions and Readmissions Project 2015-2017. [memo on the Internet] Washington, D.C.: National Quality Forum; 2016. Nov., [cited 2017 Jan 19]. Available from: http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=84046

Wigdor, A.K. and Green, B. F. Editors; Committee on the Performance of Military Personnel, National Research Council. 1991. *Performance Assessment for the Workplace, Volume I*. Washington, D.C. National Academy Press.

U.S. Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation. 2016. *Report to Congress: Social Risk Factors and Performance Under Medicare's*

Value-Based Purchasing Programs. Washington, D.C.

APPENDIX B: MEASURE EVALUATION TABLES

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
Submission
Description: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) for patients discharged from the hospital with a principal diagnosis of heart failure (HF). The outcome (readmission) is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. The target population is patients 18 and over. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals or Veterans Health Administration (VA) hospitals. **Numerator Statement**: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index HF admission. If a patient has more than one unplanned admissions (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission, because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission. **Denominator Statement**: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 18 years or older. We have explicitly tested the measure in both age groups. The cohort includes admissions for patients aged 18 years and older discharged from the hospital with either a principal discharge diagnosis of HF (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals or Veterans Health Administration (VA) hospitals. Additional details are provided in S.9 Denominator Details. **Exclusions**: The readmission measures excludes admissions: 1. Ending in discharges against medical advice Rationale: Providers did not have the opportunity to deliver full care and prepare the patient for discharge. 2. Without at least 30 days of post-discharge enrollment in FFS Medicare Rationale: The 30-day readmission outcome cannot be assessed in this group since claims data are used to determine whether a patient was readmitted. 3. Occurring within 30 days of discharge from an index admission Rationale: This exclusion ensures that no hospitalization will be considered as both a readmission and an index admission within the same measure.

NATIONAL QUALITY FORUM

Memo

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

4. With a procedure code for LVAD implantation or heart transplantation either during the index admission or in the 12 months prior to the index admission

Rationale: Patients with these procedures are a highly-selected group of patients with a different risk of the readmission outcome.

Adjustment/Stratification: Risk Adjustment Type: Statistical risk model; Risk Adjustment Methodology: Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

Type of Measure: Outcome

Data Source: Administrative claims

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING [06/08/2016-06/09/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-19; N-1** 1b. Performance Gap: **H-9**; **M-9**; **L-1**; **I-1**

<u>Rationale</u>:

- Data provided by the developer cover a total of 1,210,454 and show that heart failure readmission rates ranges from a minimum of 16% to a maximum of 32.1%.
- Hospitals serving low proportion of Dual Eligible, African-American, and patients below AHRQ SES index score of 42.7 had lower readmission rates than those with high proportions of these patients.
- The Standing Committee discussed the two updates to the measure. First, the updated measure excludes patients who have either an LVAD or a heart transplant during their indexed stay or during the year prior. The Standing Committee generally agreed that this change was an appropriate reflection of a change in clinical practice. Second, the measure had modest changes to the planned readmissions algorithm which excludes scheduled or planned readmissions from the measure.
- The noted that there is still a performance gap, with the average heart failure readmission rate over 22 percent and rates ranging from 16 percent to over 32 percent.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-7**; **M-12**; **L-0**; **I-1** 2b. Validity: **H-1**; **M-17**; **L-1**; **I-1**

<u>Rationale</u>:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
- In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

- The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.
- A total of 1,210,454 admissions over a 3-year period were examined, with 604,022 in one sample and 606,432 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples. The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.58.
- The developer demonstrated measure validity through medical record validation.
    - The HF readmission administrative model (original model specification prior to completion of the planned readmission algorithm) was validated against a medical record model with the same cohort of patients for whom hospital-level HF readmission medical record data are available.
    - A measure cohort was developed with medical record data using the inclusion/exclusion criteria and risk-adjustment strategy.
    - A sample of 64,329 patients was matched for comparison.
- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR). Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. The C-statistic is 0.63.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- SDS variables were ultimately not included in model as the developer found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
- The Standing Committee expressed concerns about published literature suggesting there was a small, but significant inverse correlation between readmissions and mortality and recommended continued monitoring.
- Overall, the Standing Committee agreed this measure met the reliability and validity criteria.

3. Feasibility: H-16; M-4; L-;0 I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c. Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- This measure is collected through administrative claims data. The Standing Committee agreed the measure would be feasible to collect and implement.

4. Usability and Use: H-3; M-15; L-2; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

- It is currently used in the Hospital Inpatient Quality Report (IQR) and Hospital Readmissions Reduction (HRRP) Programs.
- The Standing Committee noted that this measure is associated with reduction in hospital RSRR by 1.6% between 2011-2012 and 2013-2014.
- The Standing Committee agreed the measure is highly usable.

5. Related and Competing Measures

- This measure is related to #2880: Excess days in acute care (EDAC) after hospitalization for heart failure.

Standing Committee Recommendation for Endorsement: **Y-19**; **N-1**

Rationale

- The Standing Committee recognized the importance of reducing readmissions due to heart failure and the need for improved care coordination and recommended the measure for continued endorsement.

6. Public and Member Comment

- This measure received four comments. One commenter raised concerns about the inverse correlation between readmissions and mortality for heart failure. The commenter also raised concerns that the Intra-Class Correlation Coefficient (ICC) for this measure was .58. The commenter also questioned that the C-statistic for this measure was 0.63.
- Three commenters noted the need for this measure to be adjusted for SDS factors.
- One commenter raised concerns that this measure could be implemented at levels of analysis other than the one for which it is endorsed. This commenter also raised concerns that this measure competes with #0277.
- One commenter raised concerns about the relationship between declining hospital admission rates and readmissions.
- Developer response:
  - Inverse correlation between readmissions and mortality
    - The hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) and risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization have been publicly reported since June 2007 and June 2009, respectively. Yale-CORE reported the results of an examination of the correlation between the two outcomes using CMS claims data from 2005-2008 in a published study (Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA 2013; 309:587-593). The results demonstrated that the correlation, although statistically significant, is relatively low (the Pearson correlation is -0.17 with a 95% CI of -0.20 to -0.14) and only exists in the lower range of RSMRs. The much more dominant finding is that hospitals can perform well on both measures and that a relatively important share of hospitals perform above the national average or below the national average on both mortality and readmission measures. These results, which are consistent across different types of hospitals, such as teaching hospitals and rural hospitals, demonstrate that there is no systematic relationship between the two measures.
  - Intra-Class Correlation Coefficient

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

- We used the Inter-Class Correlation (ICC) method to establish the reliability of the measure score. Our approach to assessing reliability is to consider the extent to which assessments of a hospital using different but randomly selected subsets of patients produces similar measures of hospital performance. That is, we take a "test-retest" approach in which hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first, and finally comparing the agreement between the two resulting performance measures across hospitals (Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. Statistics in Medicine 2002;21:3431-3446.). This is a purposefully conservative approach to assessing reliability and traditional thresholds for acceptability do not apply to interpreting these results.
- The minimally acceptable threshold noted by AHS is not appropriate for this particular analytic approach. We have cited the more appropriate convention, which describes the ICC values as moderate (0.41-0.60) for this measure (Landis JR and Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977; 33:159-174).

- SDS adjustment
  - CMS and Yale-CORE examined heart failure readmission measure results, or hospitals' performance on this measure, using their entire patient populations including both patients with and without low SES risk variables and we found observed that hospitals had similar performance in both groups. Additionally, we examined the impact of adding patient-level risk adjustment which aims to answer the extent to which patients' SES affects measure results and found very little difference in hospital scores. We also examined risk models that included all patient comorbid conditions, both SES variables (dual eligibility and AHRQ SES Index Score) and African-American or non-African-American race, and found no change to the c-statistics compared with models that did not include SES and race variables.
- C-statistic
  - A higher C-statistic is not always better for outcome quality measures. The goal of the measures is to assess quality by estimating hospital outcome rates and accounting for important patient factors. It is not to produce the best model for predicting patient outcomes. Considering an extreme example of an outcome which is fully determined by hospital care and not at all influenced by patient risk factors of any sort, we would in that case expect to observe a C-statistic of 0.5. But if hospital quality, not patient factors, were responsible for the outcome we would still conclude that this was a good quality measure, but simply that risk adjustment was unnecessary. In the case of the readmission measures, patient factors are not particularly strong predictors of the readmission outcome and our C-statistics for readmission are consistent with those reported in the literature as appropriate for assessing quality (Kansagara D., Englander H., Salatrino A., et al. Risk Prediction Models for Hospital Readmission A Systematic Review. JAMA 2011; 306(15): 1688-1698; Bradley E, Yakusheva O, Horwitz LI, Sipsma H, Fletcher J. Identifying Patients at Increased Risk for Unplanned

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

Readmission. Medical care. 2013;51(9):761-766). A crucial additional note is that, because differences in RSRRs are intended to reflect differences in quality of care among hospitals, we purposefully do not account for any aspect of the care patients receive in our risk models. You can achieve a higher C-statistic by adding information about care received to the risk model, such as interventions but such models would provide less ability to illuminate differences in quality across hospitals as they would adjust away some of the quality signal. Similarly, we could increase C-statistics by including in-hospital complications as risk factors, but it would be inappropriate for the purpose of assessing hospital quality of care.

- Committee response:
  - The Committee has reviewed your comment and appreciates your input. Consideration of sociodemographic factors in risk adjustment models is a critical issue in measurement science. The Committee takes the concerns raised by the commenters seriously. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer.

    The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and socio-demographic risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between SDS factors and readmissions. The Committee's deliberations on the need for SDS adjustment were challenging.

    The Committee noted particular limitations for measures that were conditionally endorsed based on the need for review under the NQF trial period for SDS adjustment. The committee acknowledged that measure developers were not required to address social determinants in the original analyses required for NQF review and endorsement, which contributed to the relative lack of data to ensure robust assessment of the impact of SDS in many of the post-hoc analyses.

    The Committee reiterated that their focus was on the adjustments the developer was able to put forward at this time given the data currently available. While the adjustments put forward for these measures at this time did not reach a threshold of significance the Committee was comfortable with the Committee recognizes that risk adjustment for SDS factors is a rapidly progressing area and that more work is needed to appreciate the effects of social risk, understand the most relevant patient- and community level risk factors, collect data on these risk factors, and determine the best methods to incorporate these risk factors into performance measures.

    The Committee stressed the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards. The Committee recognized the need to ensure facilities serving vulnerable populations are not penalized unfairly while at the same time balancing concerns about worsening healthcare disparities. The Committee looks forward to continued deliberations on these issues and to reexamining these measures as better data emerges. The Committee recommends a reassessment of the availability of SDS variables and a reexamination of these measures through the NQF annual update process.

0330 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization
○ The Committee endorsed this measure for hospital-level analysis based on the testing results submitted for review. The Committee agrees that this measure should not be used for individual or group practices unless updated testing and specifications are provided to the Standing Committee to support endorsement for that use case. ○ The Committee followed NQF's guidance on measure harmonization throughout the evaluation process. Prior to the in-person meeting, the Committee received materials regarding these competing measures, and held a separate call after the in-person meeting on September 1 to discuss harmonization issues and allow the developers to answer questions from Committee members. The Committee then voted via survey to recommend both measures. The Committee considered the added value and burden of recommending both measures and agreed that the differences in measure specifications added sufficient value to offset any potential negative impact. ○ The Committee has reviewed your comment and appreciates your input. The Committee recognizes the potential for negative unintended consequences of admissions and readmissions measures and recommends careful monitoring of their implementation. Above all, the Committee agreed that use of these measures should be monitored to ensure they do not inadvertently reduce access to necessary care. The Committee noted the inverse relationship between mortality and readmission for heart failure and recognized the need for careful surveillance and balancing of these measures. The Committee also reiterated its concerns about the need to carefully balance implementation of measures addressing psychiatric readmissions to prevent the risk of higher suicidality. On the other hand, the Committee has noted the desire to understand a patient's need for any subsequent acute care after a hospitalization. In particular, the Committee recognized the need understand if patients are being seen in the Emergency Department after discharge or being placed in observation. The Committee recommends continued work to ensure that the use of readmissions measures does not result in unnecessary or avoidable use of the emergency department or observation status while ensuring that all patients have access to any necessary care. The Committee noted that a number of measures recommended for endorsement in this project could help to balance these concerns, in particular the measures addressing excess days in acute care and population-based admission measures.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-12; N-0 **Decision: Approved for continued endorsement**
8. Board of Directors Vote: Yes (December 8, 2016) **Decision: Ratified for continued endorsement**
9. Appeals **Appeal received**

NATIONAL QUALITY FORUM

Memo

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

Submission

Description: The measure estimates a hospital-level 30-day, all-cause, risk-standardized readmission rate (RSRR) for patients discharged from the hospital with either a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary diagnosis of pneumonia (including aspiration pneumonia) coded as present on admission (POA). Readmission is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission. A specified set of planned readmissions do not count as readmissions. CMS annually reports the measure for patients who are 65 years or older and are enrolled in fee-for-service (FFS) Medicare hospitalized in non-federal hospitals.

Please note this measure has been substantially updated since the last submission; as described in S.3., the cohort has been expanded. Throughout this application we refer to this measure as version 8.2.

Numerator Statement: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients 18 and older discharged from the hospital with a principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

Denominator Statement: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or over or (2) patients aged 18 years or older. We have specifically tested the measure in both age groups.

The cohort includes admissions for patients aged 18 years and older discharged from the hospital with principal discharge diagnosis of pneumonia, including aspiration pneumonia or a principal discharge diagnosis of sepsis (not severe sepsis) with a secondary discharge diagnosis of pneumonia (including aspiration pneumonia) coded as POA and no secondary discharge diagnosis of severe sepsis; and with a complete claims history for the 12 months prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals.

Additional details are provided in S.9 Denominator Details.

Exclusions: The readmission measures exclude index admissions for patients:

1. Discharged against medical advice (AMA);

2. Without at least 30 days post-discharge enrollment in FFS Medicare;

3. Admitted within 30 days of a prior index admission.

Adjustment/Stratification: Risk Adjustment Type: Statistical risk model; Risk Adjustment Methodology: Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes” (Krumholz et al., 2006).

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

NATIONAL QUALITY FORUM

Memo

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

Type of Measure: Outcome

Data Source: Administrative claims

Measure Steward: Centers for Medicare & Medicaid Services

STANDING COMMITTEE MEETING [06/08/2016-06/09/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-20; N-0**; 1b. Performance Gap: **H-12**; **M-7**; **L-1**; **I-0**

<u>Rationale</u>:

- New evidence is provided since the last endorsement maintenance review. Since its last review, this measure has been updated to include an expanded cohort to include patients with aspiration pneumonia and sepsis.
- Data provided by the developer cover a total of 1,469,277 and show that pneumonia readmission rates ranges from a minimum of 13.1% to a maximum of 24.7%.
- Hospitals serving low proportion of Dual Eligible, African-American, and patients below AHRQ SES index score of 42.7 had lower readmission rates than those with high proportions of these patients.
- The Standing Committee reviewed the two measure updates. First, the measure has an expanded cohort including patients who have a principal diagnosis of sepsis and a secondary diagnosis of pneumonia that is present on admission, and patients who have a principal diagnosis of aspiration pneumoia. Second, the measure includes the updated planned readmissions algorithm noted for Measure #0330.
- The Standing Committee agreed that the measure still has a performance gap, with rates of pneumonia readmission ranging from 13.1 percent to 24.7 percent with an average rate of 17.5 percent.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-12**; **M-8**; **L-0**; **I-0** 2b. Validity: **H-1**; **M-17**; **L-1**; **I-0**

<u>Rationale</u>:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
- In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.
- A total of 1,469,277 admissions over a 3-year period were examined, with 733,434 in one sample and 735,843 in the other randomly-selected sample. Two risk-standardized mortality rates (RSMR) were calculated for each hospital: one from each of the two separate samples.

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

The agreement between the two RSMRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.73.

- The developer tested the original version of the measure by comparing the administrative model with a medical-record based model. The results of this testing are included in the citation Krumholz, 2008. The developer notes that the claims-based measure produced results which were highly correlated with those produced through manual chart audit. (Krumholz et al., 2008; Lindenauer et al., 2011)
- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR). Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. The C-statistic is 0.63.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- SDS variables were ultimately not included in model as the developer found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
- The Standing Committee questioned whether hospitals with a larger proportion of aspiration pneumonia patients did similar to other hospitals, to which the developers noted yes. Additionally, Standing Committee members expressed concerns on the lack of data published on sensitivity and specificity for patients with sepsis and pneumonia as a secondary diagnosis in hospitals. Overall, the Standing Committee agreed this measure had sufficient reliability and validity testing to meet the criteria.

3. Feasibility: H-15; M-4; L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- This measure is collected through administrative claims data. The Standing Committee agreed the measure would be feasible to collect and implement.

4. Usability and Use: H-4; M-15; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- It is currently used in the Hospital Inpatient Quality Report (IQR) and Hospital Readmissions Reduction (HRRP) Programs.
- The Standing Committee agreed the measure is highly usable.

5. Related and Competing Measures

- This measure is related to NQF #0279: Bacterial Pneumonia Admission Rate (PQI 11) and NQF #2882: Excess days in acute care (EDAC) after hospitalization for pneumonia. The developer notes that the measures are not completely harmonized. The developer justifies the difference by noting that for outcome measures clinical coherence of the cohort takes precedence over alignment with related non-outcome measures.

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

Standing Committee Recommendation for Endorsement: **Y-18**; **N-1**

Rationale

- The Standing Committee recognized the importance of reducing readmissions due to pneumonia and the need for improved care coordination and discharge management and recommended the measure for continued endorsement.

6. Public and Member Comment

- Four comments were received on this measure. One commenter raised concerns about the expansion of the measure cohort. Two commenters expressed concerns that the measure does not include SDS factors in the risk adjustment model. One commenter noted that his measure should account for planned admissions. One commenter raised concerns about the relationship between declining admission rates and readmissions.
- Developer response:
  - Expanded cohort:
    - Several studies in the published literature have shown a rapid increase in the use of sepsis codes over recent years for patients with pneumonia, and wide variation in the use of sepsis as a principal discharge diagnosis code across hospitals (see, e.g., Sjoding MW, Iwashyna TJ, Dimick JB, et al. Gaming hospital-level pneumonia 30-day mortality and readmission measures by legitimate changes to diagnostic coding. Crit Care Med. 2015; 43(5): 989-995). Analyses conducted by Yale-CORE as a part of measure reevaluation demonstrated that expansion of the cohorts to include patients with sepsis and aspiration pneumonia appeared to mitigate or resolve bias in hospital performance related to diagnostic coding patterns. For example, our analyses confirmed findings of previous studies showing that hospital coding for sepsis was strongly associated with hospital performance on the pneumonia mortality and readmission measures. These findings suggested that at some hospitals where the sepsis code was used more frequently, patients who met the diagnostic definition for sepsis and also had pneumonia were being excluded from the measure. This pattern of excluding potentially sicker pneumonia patients from the measure cohort was biasing the measure in favor of hospitals with high rates of sepsis coding. We found that this issue was resolved by the addition of the sepsis patients to the measure. Patients with pneumonia severe enough to be admitted to the hospital frequently meet criteria for sepsis and should be a part of the measure cohort. However, we do not include patients with severe sepsis.
    - Similar patterns were found in aspiration pneumonia. Hospitals used aspiration pneumonia as a principal discharge diagnosis code to varying degrees and therefore not including these patients in the measure could lead to differential exclusions across hospitals. Additionally, there is no commonly accepted definition or gold standard diagnostic test to identify aspiration pneumonia. This is a subset of bacterial pneumonia which is diagnosed clinically, often subjectively based on patient's risk factors, such as age and frailty. The treatment of patients who receive a diagnosis for aspiration pneumonia is carried out by the same care teams and using similar approaches as patients with other types of pneumonia. Additionally,

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

the prospective payment system creates strong incentives for hospitals to make a diagnosis of aspiration pneumonia because it changes the Diagnostic Related Group (DRG) from simple pneumonia (MS-DRG 177-179) to a higher reimbursement DRG for respiratory infections and inflammations (MS-DRG 193-195). Variation across hospitals in the application of the aspiration pneumonia code has the potential to bias outcomes estimated across hospitals when calculated for the mortality and readmission measures, in the same way that variation in sepsis coding has been shown to introduce bias in the pneumonia measures.

- Our findings argued for the need to broaden the cohort to capture the full spectrum of disease presentation and to reduce potential bias. The rationale for expansion was based on the intent, to include the full spectrum of patients admitted for the treatment of pneumonia and to prevent bias based on different coding patterns.

- Planned readmission
  - The CMS readmission measures do not consider planned readmissions as part of the readmission outcome. Generally speaking, planned readmissions are not a signal of quality of care. Therefore, CMS has worked with experts in the medical community as well as other stakeholders to carefully identify procedures and treatments that should be considered "planned," and thus not considered in the readmission outcome. Starting with the 2013 public reporting, the measures identify planned readmissions by using an expanded algorithm, which is a set of criteria for classifying readmissions as planned using Medicare claims. This algorithm identifies admissions that are typically planned and may occur within 30 days of discharge from the hospital. The algorithm is based on three principles: • A few specific, limited types of care are always considered planned (transplant surgery, maintenance chemotherapy/radiotherapy/ immunotherapy, rehabilitation);• Otherwise, a planned readmission is defined as a non-acute readmission for a scheduled procedure; and • Admissions for acute illness or for complications of care are never planned. CMS conducted a validation study of the planned readmission algorithm using medical record data from 634 medical records at seven hospitals. For the 2016 public reporting, Version 4.0 of the algorithm includes modifications to enhance the accuracy of the algorithm based on the study findings. These changes improve the accuracy of the algorithm by decreasing the number of readmissions that the algorithm mistakenly designates as planned or unplanned. This involved the removal of five procedure categories and the addition one procedure category to the list of potentially planned procedure that disqualify readmissions from the measure outcome.
  - For the details of the planned readmission algorithm as applied to the pneumonia measure, refer to Appendix E of the 2016 Condition-Specific Readmission Measures Updates and Specifications Report available on QualityNet at: (www.qualitynet.org) > Hospitals – Inpatient > Claims-Based Measures > Readmission Measures > 2016 AMI, HF, Pneumonia, COPD, and Stroke Readmission Measures Updates and Specifications Report (also available at

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890567694&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DCondSpecific_Rdmsn_Rpt_2016.pdf&blobcol=urldata&blobtable=MungoBlobs).

- SDS adjustment
  - CMS agrees that patients' socioeconomic status (SES) affects health and health outcomes in important ways. In the conceptual model presented to the Committee, we explain that many patients with low SES indicators may have poorer health status at the start of an index admission that increases their risk of readmission. The decrease in the strength of the association between SES variables and the readmission outcome when we added patients' comorbidities to the risk model supports this proposed mechanism. Additionally, the results presented showed that the effect of SES variables on readmission rates in the multi-variate or fully adjusted model was small but significant. However, inclusion of these variables did not change hospitals risk-standardized readmission rates or their performance on the measures. Yale-CORE remains committed to examining alternative solutions that better reflect the balance of hospital- and patient-level influences on hospital outcome measures and to considering appropriate ways to incorporate community factors into the outcomes measures.

- Committee response:
  - Consideration of sociodemographic factors in risk adjustment models is a critical issue in measurement science. The Committee takes the concerns raised by the commenters seriously. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer.The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and socio-demographic risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between SDS factors and readmissions. The Committee's deliberations on the need for SDS adjustment were challenging.
    The Committee noted particular limitations for measures that were conditionally endorsed based on the need for review under the NQF trial period for SDS adjustment. The committee acknowledged that measure developers were not required to address social determinants in the original analyses required for NQF review and endorsement, which contributed to the relative lack of data to ensure robust assessment of the impact of SDS in many of the post-hoc analyses.

    The Committee reiterated that their focus was on the adjustments the developer was able to put forward at this time given the data currently available. While the adjustments put forward for these measures at this time did not reach a threshold of significance the Committee was comfortable with the Committee recognizes that risk adjustment for SDS factors is a rapidly progressing area and that more work is needed to appreciate the effects of social risk, understand the most relevant patient- and community level risk factors, collect data on these risk factors, and determine the best methods to incorporate these risk factors into performance measures.

0506 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization
○ The Committee stressed the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards. The Committee recognized the need to ensure facilities serving vulnerable populations are not penalized unfairly while at the same time balancing concerns about worsening healthcare disparities. The Committee looks forward to continued deliberations on these issues and to reexamining these measures as better data emerges. The Committee recommends a reassessment of the availability of SDS variables and a reexamination of these measures through the NQF annual update process.The Committee recognizes the potential for negative unintended consequences of admissions and readmissions measures and recommends careful monitoring of their implementation. Above all, the Committee agreed that use of these measures should be monitored to ensure they do not inadvertently reduce access to necessary care. The Committee noted the inverse relationship between mortality and readmission for heart failure and recognized the need for careful surveillance and balancing of these measures. The Committee also reiterated its concerns about the need to carefully balance implementation of measures addressing psychiatric readmissions to prevent the risk of higher suicidality. On the other hand, the Committee has noted the desire to understand a patient's need for any subsequent acute care after a hospitalization. In particular, the Committee recognized the need understand if patients are being seen in the Emergency Department after discharge or being placed in observation. The Committee recommends continued work to ensure that the use of readmissions measures does not result in unnecessary or avoidable use of the emergency department or observation status while ensuring that all patients have access to any necessary care. The Committee noted that a number of measures recommended for endorsement in this project could help to balance these concerns, in particular the measures addressing excess days in acute care and population-based admission measures.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-12; N-0 **Decision: Approved for continued endorsement**
8. Board of Directors Vote: Yes (December 8, 2016) **Decision: Ratified for continued endorsement**
9. Appeals **Appeal Received**

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
Submission
Description: The measure estimates a hospital-level risk-standardized readmission rate (RSRR) of unplanned, all-cause readmission after admission for any eligible condition within 30 days of hospital discharge. The measure reports a single summary RSRR, derived from the volume-weighted results of five different models, one for each of the following specialty cohorts based on groups of discharge

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

condition categories or procedure categories: surgery/gynecology; general medicine; cardiorespiratory; cardiovascular; and neurology, each of which will be described in greater detail below. The measure also indicates the hospital-level standardized risk ratios (SRR) for each of these five specialty cohorts. The outcome is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals.

Numerator Statement: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from an eligible index admission. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission.

Denominator Statement: The measure includes admissions for Medicare beneficiaries who are 65 years and older and are discharged from all non-federal, acute care inpatient US hospitals (including territories) with a complete claims history for the 12 months prior to admission.

Additional details are provided in S.9 Denominator Details.

Exclusions: The measure excludes index admissions for patients:

1. Admitted to Prospective Payment System (PPS)-exempt cancer hospitals;
2. Without at least 30 days post-discharge enrollment in FFS Medicare;
3. Discharged against medical advice (AMA);
4. Admitted for primary psychiatric diagnoses;
5. Admitted for rehabilitation; or
6. Admitted for medical treatment of cancer.

Adjustment/Stratification: Risk Adjustment Type: Statistical risk model; Risk Adjustment Methodology: Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

Type of Measure: Outcome

Data Source: Administrative claims

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING [06/08/2016-06/09/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Unchanged – no vote** 1b. Performance Gap: **H-3**; **M-15**; **L-1**; **I-0**

<u>Rationale</u>:

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

- The developer stated that there are no updates to the evidence since the last submission, so the Standing Committee agreed that there was no need for a repeat discussion or vote on evidence.
- Data provided by the developer cover a total of 22,000,000 admissions and show that readmission rates ranges from a minimum of 11.4% to a maximum of 20.1%.
- Hospitals serving low proportion of Dual Eligible, African-American, and patients below AHRQ SES index score of 45 had lower readmission rates than those with high proportions of these patients.
- The Standing Committee agreed that there continues to be a performance gap, with all cause readmission rates ranging from 11.4 percent to 20.1 percent with an average of 15.4 percent.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-8**; **M-11**; **L-0**; **I-0** 2b. Validity: **H-8**; **M-11**; **L-0**; **I-0**

<u>Rationale</u>:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
- In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method. This is generally considered an appropriate method of testing reliability.
- A total of 6,843,808 admissions in the 2015 publicly reported measure, with 3,420,728 in one sample and 3,423,080 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples. The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.80.
- The developer demonstrated measure validity through prior validity testing done on their other claims-based measures, through the use of established measure development guidelines, and examination of content validity by comparing hospital performance with that on other quality measures.
- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR). Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity.
- C-statistic for each cohort:
  - Medicine cohort: 0.643
  - Surgical cohort: 0.675
  - Cardiorespiratory cohort: 0.636
  - Cardiovascular cohort: 0.658

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
○ Neurology cohort: 0.622 • The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources. • SDS variables were ultimately not included in model as the developer found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance. • The Standing Committee raised concerns that merging multiple cohorts into one group may mask the individual variance properties of the individual cohorts. • The Standing Committee expressed that the modeling was laid out very explicitly and well-specified and generally agreed that the measure had sufficient reliability and validity testing to meet the reliability and validity criteria.
3. Feasibility: H-17; M-2; L-0; I-0 *(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)* Rationale: • This measure is collected through administrative claims data. The Standing Committee agreed the measure would be feasible to collect and implement.
4. Usability and Use: H-7; M-12; L-0; I-0 *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: • It is currently used in the Hospital Inpatient Quality Report (IQR) Program. • The Standing Committee agreed the measure is highly usable.
5. Related and Competing Measures • This measure is related to NQF # 1768: Plan All-Cause Readmissions (PCR). This measure and the NCQA Plan All-Cause Readmissions (PCR) Measure #1768 are related measures, but are not competing because they don't have the same measure focus and same target population. Each of these measures has different specifications. In addition, both have been previously harmonized to the extent possible under the guidance of the National Quality Forum Standing Committee in 2011.
Standing Committee Recommendation for Endorsement: **Y-18**; **N-1** Rationale • The Standing Committee recognized the importance of reducing readmissions and the need for improved care coordination and discharge management and recommended the measure for continued endorsement.
6. Public and Member Comment • Eight comments were received on this measure, including a sign-on letter from 28 physician societies. Two commenters expressed their support for endorsement of this measure. • A number of commenters raised concerns that NQF #1789 Hospital-wide all-cause unplanned readmission measure is being used at the clinician level of analysis in the Physician Value-Based Payment Modifier program and is proposed to be used in the Merit-Based Incentive Payment System in a similar way. These commenters expressed concern that testing at this level of analysis was not provided to the Standing Committee for review.

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

- Two commenters raised concerns that this measure does not include SDS factors in the risk adjustment model.
- One commenter expressed concerns that trauma is not excluded from the measure.
- Developer Response:
  - SDS adjustment
    - CMS agrees that patients' socioeconomic status (SES) effects health and health outcomes in important ways. In the conceptual model presented to the Committee, we explain that many patients with low SES indicators may have poorer health status at the start of an index admission that increases their risk of readmission. The decrease in the strength of the association between SES variables and the readmission outcome when we added patients' comorbidities to the risk model supports this proposed mechanism. Additionally, the results presented showed that the effect of SES variables on readmission rates in the multi-variate or fully adjusted model was small but significant. However, inclusion of these variables did not change hospitals risk-standardized readmission rates or their performance on the measures. We explained that the remaining small effect of SES in the risk models could be a hospital-level effect, if patients with low SES indicators more often receive care at lower quality hospitals. Alternatively, it could be a patient-level effect, if patients have other unmeasured factors that increase their risk of readmission that are beyond the hospitals' control or if they receive inappropriate care from hospitals due to bias or discrimination. The results of the decomposition analyses we presented to the Committee confirmed that most of the small residual effect of SES variables on readmission rates is a hospital-level effect, suggesting that it is due to the clustering of patients with low SES indicators and low quality hospitals. Therefore, we concluded that the evidence did not support including SES variables in the measures risk models. We also note that the lack of any change in hospitals performance with inclusion of individual SES risk variables also held true when all SES variables were added to the fully adjusted model together. Yale-CORE remains committed to examining alternative solutions that better reflect the balance of hospital- and patient-level influences on hospital outcome measures and to considering appropriate ways to incorporate community factors into the outcomes measures.
  - Relationship between admission and readmission rates
    - In a recent study published in Health Affairs, Dharmarajan and colleagues (Dharmarajan K, Qin L, Lin ZQ, et al. Declining Admission Rates And Thirty-Day Readmission Rates Positively Associated Even Though Patients Grew Sicker Over Time. Health Affairs 2016; 35(7): 1294-1302) explore the relationship between admission and readmission rates. Using national data on Medicare fee-for-service beneficiaries from 2010 to 2013, the study shows that communities with a decline in admission rates also had a decline in readmission rates despite the fact that hospitalized patients were sicker. This association suggests that reducing admission rates does not necessarily lead to higher readmission rates. From a policy perspective, both outcomes might be pursued simultaneously.
  - 2015 Physician Fee Schedule Proposed Rule

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

- Questions about application of this measure beyond the hospital setting is beyond the scope of what the developer was asked to examine and consider for measure endorsement maintenance.

- Physician level reliability
  - This comment is out of scope for the hospital measures.
- Trauma
  - The CMS readmission measures assess all-cause readmissions; that is, they consider unplanned readmissions for any reason, not only those that are due to the same or a "related" condition.
  - There are several reasons for measuring all-cause readmissions. First, from the patient perspective, an unplanned readmission is disruptive and costly regardless of cause. Second, restricting the measure outcomes to those readmissions that seem to be directly related to the initial hospitalization may make the measures susceptible to gaming through changes in coding practices. Although most hospitals would not engage in such practices, CMS wants to eliminate any incentive for hospitals to change coding practices in an effort to prevent readmissions from being captured in their readmission measure results. Third, an apparently unrelated readmission may represent a complication related to the underlying condition. For example, a patient with heart failure who develops a hospital-acquired infection may later be readmitted due to that infection. It would be inappropriate to consider this readmission as unrelated to the care the patient received for heart failure. Finally, hospitals can act to reduce readmissions from all causes. While CMS does not presume that every readmission is preventable, measuring all-cause readmission incentivizes hospitals to evaluate the full range of factors that increase patients' risk for unplanned readmissions. For example, unclear discharge instructions, poor communication with post-acute care providers, and inadequate follow-up are factors that typically increase the risk for an unplanned readmission.
  - Although measuring all-cause readmissions will include some patients whose readmission may be unrelated to their care (for example, a casualty in a motor vehicle accident), such events should occur randomly across hospitals and therefore will not affect results on measures that assess relative performance.
  - Note that planned readmissions do not count as readmissions in the 30-day readmission measures. For the details of the planned readmission algorithm as applied to the HWR measures, refer to Appendix E of the 2016 Hospital-wide Readmission Measure Updates and Specifications Report available on QualityNet at: (www.qualitynet.org) > Hospitals – Inpatient > Claims-Based Measures > Readmission Measures > Measure Methodology > Hospital-Wide Readmission Measure Methodology Report (also available at http://www.qualitynet.org/dcs/BlobServer?blobkey=id&blobnocache=true&blobwhere=1228890434757&blobheader=multipart%2Foctet-stream&blobheadername1=Content-Disposition&blobheadervalue1=attachment%3Bfilename%3DDryRun_HWR_TechReport_081012%2C0.pdf&blobcol=urldata&blobtable=MungoBlobs). Proposed Committee Response: Thank you for your comment. The Committee endorsed this measure for hospital-level analysis based on the

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

testing results submitted for review. The Committee agrees that this measure should not be used for individual or group practices unless updated testing and specifications are provided to the Standing Committee to support endorsement for that use case. The Committee encourages the measure developer to bring additional testing results for alternative use cases to NQF for multistakeholder review.

- Committee response:
  - The Committee endorsed this measure for hospital-level analysis based on the testing results submitted for review. The Committee agrees that this measure should not be used for individual or group practices unless updated testing and specifications are provided to the Standing Committee to support endorsement for that use case. The Committee encourages the measure developer to bring additional testing results for alternative use cases to NQF for multistakeholder review.
  - The Committee recognizes the potential for negative unintended consequences of these measures and recommends careful monitoring of their implementation.
  - Consideration of sociodemographic factors in risk adjustment models is a critical issue in measurement science. The Committee takes the concerns raised by the commenters seriously. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer.

    The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and socio-demographic risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between SDS factors and readmissions. The Committee's deliberations on the need for SDS adjustment were challenging.

    The Committee noted particular limitations for measures that were conditionally endorsed based on the need for review under the NQF trial period for SDS adjustment. The committee acknowledged that measure developers were not required to address social determinants in the original analyses required for NQF review and endorsement, which contributed to the relative lack of data to ensure robust assessment of the impact of SDS in many of the post-hoc analyses.

    The Committee reiterated that their focus was on the adjustments the developer was able to put forward at this time given the data currently available. While the adjustments put forward for these measures at this time did not reach a threshold of significance the Committee was comfortable with the Committee recognizes that risk adjustment for SDS factors is a rapidly progressing area and that more work is needed to appreciate the effects of social risk, understand the most relevant patient- and community level risk factors, collect data on these risk factors, and determine the best methods to incorporate these risk factors into performance measures.
  - The Committee stressed the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards. The Committee recognized the need to ensure facilities serving vulnerable populations are not penalized unfairly while at the same time balancing concerns about worsening healthcare disparities. The Committee looks forward to continued deliberations on these issues and to reexamining these measures as better data emerges. The Committee recommends a

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)
reassessment of the availability of SDS variables and a reexamination of these measures through the NQF annual update process.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-12; N-0 **Decision: Approved for continued endorsement**
8. Board of Directors Vote: Yes (December 8, 2016) **Decision: Ratified for continued endorsement**
9. Appeals **Appeal received**

1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
Submission
Description: The measure estimates a hospital-level 30-day, all-cause, risk-standardized readmission rate (RSRR) for patients discharged from the hospital with either a principal discharge diagnosis of COPD or a principal discharge diagnosis of respiratory failure with a secondary diagnosis of acute exacerbation of COPD. The outcome (readmission) is defined as unplanned readmission for any cause within 30 days of the discharge date for the index admission (the admission included in the measure cohort). A specified set of planned readmissions do not count in the readmission outcome. CMS annually reports the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and hospitalized in non-federal hospitals. **Numerator Statement**: The outcome for this measure is 30-day readmission. We define readmission as an inpatient admission for any cause, with the exception of certain planned readmissions, within 30 days from the date of discharge from the index admission for patients discharged from the hospital with a principal discharge diagnosis of COPD or principal discharge diagnosis of respiratory failure with a secondary discharge diagnosis of acute exacerbation of COPD. If a patient has more than one unplanned admission (for any reason) within 30 days after discharge from the index admission, only the first one is counted as a readmission. The measure looks for a dichotomous yes or no outcome of whether each admitted patient has an unplanned readmission within 30 days. However, if the first readmission after discharge is considered planned, any subsequent unplanned readmission is not counted as an outcome for that index admission because the unplanned readmission could be related to care provided during the intervening planned readmission rather than during the index admission. **Denominator Statement**: This claims-based measure can be used in either of two patient cohorts: (1) patients aged 65 years or older or (2) patients aged 40 years or older. We have explicitly tested the measure in both age groups. The cohort includes admissions for patients discharged from the hospital with either a principal discharge diagnosis of COPD (see codes below) OR a principal discharge diagnosis of respiratory failure (see codes below) with a secondary discharge diagnosis of acute exacerbation of COPD (see codes below) and with a complete claims history for the 12 months prior to admission. The measure is currently publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals. Additional details are provided in S.9 Denominator Details.

NATIONAL QUALITY FORUM

Memo

1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

Exclusions: The readmission measures exclude index admissions for patients:

1. Without at least 30 days post-discharge enrollment in FFS Medicare.
2. Discharged against medical advice (AMA);
3. Admitted within 30 days of a prior index admission.

Adjustment/Stratification: Risk Adjustment Type: Statistical risk model; Risk Adjustment Methodology: Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

Type of Measure: Outcome

Data Source: Administrative claims

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING [06/08/2016-06/09/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **unchanged – no vote**; 1b. Performance Gap: **H-7**; **M-13**; **L-0**; **I-0**

<u>Rationale</u>:

- The developer states that there are no updates to the evidence since the last submission, so the Standing Committee agreed that there was no need for a repeat discussion on evidence.
- Data provided by the developer cover a total of 925,315 admissions and show that COP readmission rates ranges from a minimum of 15.5% to a maximum of 26.6%.
- Hospitals serving low proportion of Dual Eligible, African-American, and patients below AHRQ SES index score of 45 had lower readmission rates than those with high proportions of these patients.
- The Standing Committee agreed that the measure continues to have a performance gap with readmission rates for COPD ranging from 15.5 percent to 26.6 percent and an average of 20.2 percent.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-5**; **M-14**; **L-0**; **I-0** 2b. Validity: **H-8**; **M-12**; **L-0**; **I-0**

<u>Rationale</u>:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS.
- In addition, the developer compared frequencies and odds ratios of variables from their risk model across three years of data in order to assess the consistency of those variables over time.
- The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-

1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

retest" approach; it may also be called a "split-half" method. This is generally considered to be an appropriate method of testing reliability.

- A total of 925,315 admissions over a 3-year period were examined, with 461,505 in one sample and 463,810 in the other randomly-selected sample. Two risk-standardized readmission rates (RSRR) were calculated for each hospital: one from each of the two separate samples. The agreement between the two RSRRs for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.48.
- The developer demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
- This measure employs a hierarchical logistic regression model (a form of hierarchical generalized linear model [HGLM]) to create a hospital level 30-day risk-standardized readmission rate (RSRR). Variables considered for inclusion in the model were patient-level risk-adjustors that are expected to be predictive of readmission, based on empirical analysis, prior literature, and clinical judgment, including demographic factors (age, sex) and indicators of comorbidity and disease severity. The C-statistic is 0.64.
- The developers also considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- SDS variables were ultimately not included in model as the developer found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.
- While there was discussion about the modest results of the reliability testing and the use of hierachical logistical modeling, the Standing Committee agreed that the measure met the reliability and validity criteria for NQF endorsement.

3. Feasibility: H-15 M-5 L-0; I-0

(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented)

Rationale:

- This measure is collected through administrative claims data. The Standing Committee agreed the measure would be feasible to collect and implement.

4. Usability and Use: H-9; M-11; L-0; I-0

(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)

Rationale:

- It is currently used in the Hospital Inpatient Quality Report (IQR) and Hospital Readmissions Reduction (HRRP) Programs.
- The Standing Committee noted that there are no unintended consequences for the measure, but had a few concerns regarding the advantages and disadvantages of the hierarchical approach how closely the predictive rate reflects hospital performance.
- Overall, the Standing Committee felt the measure met the NQF criteria for usability and use.

5. Related and Competing Measures

- This measure is related to NQF #0275: Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate (PQI 5).

NATIONAL QUALITY FORUM

Memo

1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

Standing Committee Recommendation for Endorsement: **Y-18**; **N-2**

Rationale

- The Standing Committee recognized the importance of reducing readmissions due to COPD and the need for improved care coordination and discharge management and recommended the measure for continued endorsement.

6. Public and Member Comment

- Three comments were received on this measure. One commenter submitted a comment in support of recommending the measure for endorsement.
- One commenter raised concerns about the level of reliability for this measure, saying the Intra-Class Correlation Coefficient (ICC) of 0.48 was low and an ICC of 0.60 should be the threshold.
- One commenter raised concerns about the potential unintended consequences of endorsing the measure, and the unknown number of truly preventable readmissions.
- Committee Response:
  - . While the measure that was submitted to NQF has an Intra-Class Correlation Coefficient below 0.60, the Committee believes it represents an acceptable benchmark for reliability for measurement of readmissions following a hospitalization for COPD. The Committee concluded that developers' current approach to risk-adjustment and exclusions met the Scientific Acceptability criteria, and were satisfied with the measure's reliability.
  - The Committee recognizes the potential for negative unintended consequences of admissions and readmissions measures and recommends careful monitoring of their implementation. Above all, the Committee agreed that use of these measures should be monitored to ensure they do not inadvertently reduce access to necessary care. The Committee noted the inverse relationship between mortality and readmission for heart failure and recognized the need for careful surveillance and balancing of these measures. The Committee also reiterated its concerns about the need to carefully balance implementation of measures addressing psychiatric readmissions to prevent the risk of higher suicidality.
  - On the other hand, the Committee has noted the desire to understand a patient's need for any subsequent acute care after a hospitalization. In particular, the Committee recognized the need understand if patients are being seen in the Emergency Department after discharge or being placed in observation. The Committee recommends continued work to ensure that the use of readmissions measures does not result in unnecessary or avoidable use of the emergency department or observation status while ensuring that all patients have access to any necessary care. The Committee noted that a number of measures recommended for endorsement in this project could help to balance these concerns, in particular the measures addressing excess days in acute care and population-based admission measures.

7. Consensus Standards Approval Committee (CSAC) Vote: Y-12; N-0

Decision: Approved for continued endorsement

8. Board of Directors Vote: Yes (December 8, 2016)

Decision: Ratified for continued endorsement

9. Appeals

NATIONAL QUALITY FORUM

Memo

1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization
Appeal received

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
Submission
Description: This measure assesses days spent in acute care within 30 days of discharge from an inpatient hospitalization for acute myocardial infarction (AMI) to provide a patient-centered assessment of the post-discharge period. This measure is intended to capture the quality of care transitions provided to discharged patients hospitalized with AMI by collectively measuring a set of adverse acute care outcomes that can occur post-discharge: emergency department (ED) visits, observation stays, and unplanned readmissions at any time during the 30 days post-discharge. In order to aggregate all three events, we measure each in terms of days. In 2016, CMS will begin annual reporting of the measure for patients who are 65 years or older, are enrolled in fee-for-service (FFS) Medicare, and are hospitalized in non-federal hospitals. **Numerator Statement**: The outcome of the measure is a count of the number of days the patient spends in acute care within 30 days of discharge. We define days in acute care as days spent in an ED, admitted to an observation unit, or admitted as an unplanned readmission for any cause within 30 days from the date of discharge from the index AMI hospitalization. Each ED treat-and-release visit is counted as one half-day (0.5 days). Observation stays are recorded in terms of hours and are rounded up to the nearest half-day. Each readmission day is counted as one full day (1 day). We count all eligible outcomes occurring in the 30-day period, even if they are repeat occurrences. **Denominator Statement**: The target population for this measure is Medicare FFS beneficiaries aged 65 years and older hospitalized at non-federal acute care hospitals for AMI. The cohort includes admissions for patients discharged from the hospital with a principal discharge diagnosis of AMI (see codes below in S.9) and with continuous 12 months Medicare enrollment prior to admission. The measure will be publicly reported by CMS for those patients 65 years and older who are Medicare FFS beneficiaries admitted to non-federal hospitals. Additional details are provided n S.9 Denominator Details. **Exclusions**: The measure excludes index admissions for patients: 1. Without at least 30 days post-discharge enrollment in FFS Medicare; 2. Discharged against medical advice (AMA); 3. Admitted within 30 days of a prior index discharge; 4. Admitted and then discharged on the same day (because it is unlikely these are clinically significant AMIs). **Adjustment/Stratification**: Risk Adjustment Type: Statistical risk model; Risk Adjustment Methodology: Our approach to risk adjustment is tailored to and appropriate for a publicly reported outcome

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

measure, as articulated in the American Heart Association (AHA) Scientific Statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes" (Krumholz et al., 2006).

Level of Analysis: Facility

Setting of Care: Hospital/Acute Care Facility

Type of Measure: Outcome

Data Source: Administrative claims

Measure Steward: Centers for Medicare & Medicaid Services (CMS)

STANDING COMMITTEE MEETING [06/08/2016-06/09/2016]

1. Importance to Measure and Report: <u>The measure meets the Importance criteria</u>

(1a. Evidence, 1b. Performance Gap)

1a. Evidence: **Y-18; N-0**; 1b. Performance Gap: **H-12**; **M-6**; **L-0**; **I-0**

<u>Rationale</u>:

- The developer cites that "the increasing use of ED visits and observation stays has raised concerns that current readmission measures do not capture the full range of unplanned acute care in the post-discharge period" (Vashi et al., 2013; Rising et al., 2012; Feng et al., 2012).
- Additionally, the developer notes that "observation stays can occur in many different parts of the hospital, including dedicated treatment rooms, the ED, or inpatient units. In particular, there is concern that high use of observation stays could in some cases replace readmissions, and that hospitals with high rates of observation stays in the post-discharge period may therefore have low readmission rates that do not accurately reflect the quality of care (Vashi et al., 2013)."
- Data provided by the developer cover a total of 232,954 discharges and show that AMI readmission rates range from a minimum of -54 to a maximum of 170.
- Similar to NQF #2880, the Standing Committee agreed that the measure has a significant performance gap with the 10th percentile -23 days to the 90th percentile at 46 days among hospitals. The Standing Committee agreed that the measure is important to measure and report.

2. Scientific Acceptability of Measure Properties: <u>The measure meets the Scientific Acceptability criteria</u>

(2a. Reliability - precise specifications, testing; 2b. Validity - testing, threats to validity)

2a. Reliability: **H-0**; **M-18**; **L-0**; **I-0** 2b. Validity: **H-1**; **M-17**; **L-0**; **I-0**

<u>Rationale</u>:

- With regard to data element reliability, the developer notes that the measure has been developed to avoid the use of claims data elements that are thought to be coded inconsistently across hospitals or providers, instead using fields that are consequential for payment and which are audited by CMS. Additionally, the developer used the final risk-adjustment variables in the existing, NQF-endorsed measure of hospital-level risk-standardized readmission rates following AMI (NQF #0505).
- The developer's approach to assessing score-level reliability was to consider the extent to which assessments of a hospital using different but randomly-selected subsets of patients produce similar measures of hospital performance. The developers refer to this as a "test-retest" approach; it may also be called a "split-half" method.
- For test-retest reliability, the developer calculated the EDAC for each hospital using first the development sample, then the validation sample. Thus, each hospital twice was measured twice, each time using an entirely distinct set of patients. The developer states that the extent

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

to which the calculated measures of these two subsets agree is evidence that the measure is assessing an attribute of the hospital, not of the patients. As a metric of agreement, the developer calculated the intra-class correlation coefficient (ICC) as defined by ICC[2,1] by Shrout and Fleiss (1979) and assessed the values according to conventional standards (Landis and Koch, 1977).

- A total of 496,716 admissions were examined, with 248,358 in each sample. The agreement between the two EDAC values for each hospital (as measured by an intra-class correlation coefficient (ICC)) was 0.54.
- The developer demonstrated measure validity through prior validity testing done on their claims-based measures, through use of established measure development guidelines, and by systematic assessment of measure face validity by a Technical Expert Panel (TEP).
- The measure employs a hierarchical generalized linear model [HGLM]) that consists of two parts, a logit model and a truncated Poisson model. The two-part logit/Poisson model (often called a "hurdle" model) assumes that the outcome results from two related processes: an initial dichotomous event – that a patient has at least one acute care event – which is modeled as the logit of the probability of the event, and for patients with an event (those which clear the "hurdle"), the number of days, which is modeled as a Poisson process. The outcome, number of days, is a half-integer count variable (because ED visits count as 0.5 days).
- The developers considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources.
- The developers state that both the patient-level and hospital-level dual eligible and race effects were significant in the logistic part of the AMI EDAC model, but only the hospital-level effect was significant in the Poisson part of the model. This indicates that a) both the patient- and hospital-level dual eligible and race effects are associated with an increased risk of acute care but b) only the hospital-level effect is associated with the expected duration of that care. The developers note that if the dual eligible or race are used in the model to adjust for patient-level differences, then some of the differences between hospitals would also be adjusted for, potentially obscuring a signal of hospital quality.
- The developers state that given these findings and complex pathways that could explain any relationship between SDS and readmission, they did not incorporate SDS variables into the measure
- The Standing Committee had moderate certainty that the measure scores are reliable and valid with an intraclass correlation coefficient of 0.54, and a correlation with readmissions of 0.61.
- For the logit model of zero versus non-zero days, which includes all patients in the cohort, the developers calculated the c-statistic.
    - C-statistic for logit part of model: 0.60
- For the Poisson model of non-zero days, which includes only patients with some acute care, the developers calculated the deviance R2. The deviance R2 is computed from the difference in the log-likelihoods between the final model and an empty model (no covariates) attributed to each observation, averaged over all observations.
    - Deviance R2 for truncated Poisson part of model: 0.040 (4.0%)
- Standing Committee members expressed that the observed to predicted graph on this measure was better than the heart failure measure #2880.
- The Standing Committee agreed this measure met the reliability and validity criteria.

3. Feasibility: H-15; M-3; L-0; I-0

NATIONAL QUALITY FORUM

Memo

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
(3a. Clinical data generated during care delivery; 3b. Electronic sources; 3c.Susceptibility to inaccuracies/ unintended consequences identified 3d. Data collection strategy can be implemented) Rationale: • This measure is collected through administrative claims data. The Standing Committee agreed the measure would be feasible to collect and implement.
4. Usability and Use: H-4; M-14; L-0; I-0 *(Used and useful to the intended audiences for 4a. Accountability and Transparency; 4b. Improvement; and 4c. Benefits outweigh evidence of unintended consequences)* Rationale: • This measure in not currently publicly reported, but was finalized for use in CMS' Hospital Inpatient Quality Report (IQR) program starting in FY 2018.
5. Related and Competing Measures • This measure is related to NQF #0505: Hospital 30-day all-cause risk-standardized readmission rate (RSRR) following acute myocardial infarction (AMI) hospitalization. The developers note that both measures are harmonized.
Standing Committee Recommendation for Endorsement: **Y-18**; **N-0** Rationale • The Standing Committee recognized the importance of reducing excess days in acute care due to acute myocardial infarction. The Standing Committee agreed that this measure is an important contribution to performance measurement as it captures the potential unintended negative consequences of increased ED use and observation stays when measuring readmissions. Standing Committee members emphasized that the developers should communicate the differences between these measures and the readmissions measures so there is no confusion, since the reporting format is not as consistent with the methods used in the past for readmissions measures.
6. Public and Member Comment • This measure received four comments. One comment was in support of recommending the measure for endorsement. • One commenter was concerned about the inconsistency between the level of analysis and level of implementation of the measure. The same commenter also raised concerns about the measure developer's decision not to include sociodemographic factors in the risk adjustment model. • One commenter raised concerns about the level of reliability for this measure, saying the Intra-Class Correlation Coefficient (ICC) of 0.48 was low and an ICC of 0.60 should be the threshold. • One commenter raised concerns about the intent of the measure and the utility of a measure that broadly defines acute care. The same commenter was concerned about the overlap of this measure and NQF #0505. • Committee Response: ○ The Committee endorsed this measure for facility-level analysis based on the testing results submitted for review. The Committee agrees that this measure should not be used for individual or group practices unless updated testing and specifications are provided to the Standing Committee to support endorsement for that use case. The Committee encourages the measure developer to bring additional testing results for alternative use cases to NQF for multistakeholder review.

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

Consideration of sociodemographic factors in risk adjustment models is a critical issue in measurement science. The Committee takes the concerns raised by the commenters seriously. The Committee was charged with evaluating the measure specifications and testing submitted on the measure as developed by the measure developer.

The Committee recognizes that there continues to be limitations in the available data elements to capture unmeasured clinical and socio-demographic risk. Given the constraints on the current data elements available, the Committee relied on the methods used by the measure developers to test the conceptual and empirical relationship between SDS factors and readmissions. The Committee's deliberations on the need for SDS adjustment were challenging.

The Committee noted particular limitations for measures that were conditionally endorsed based on the need for review under the NQF trial period for SDS adjustment. The committee acknowledged that measure developers were not required to address social determinants in the original analyses required for NQF review and endorsement, which contributed to the relative lack of data to ensure robust assessment of the impact of SDS in many of the post-hoc analyses.

The Committee reiterated that their focus was on the adjustments the developer was able to put forward at this time given the data currently available. While the adjustments put forward for these measures at this time did not reach a threshold of significance the Committee was comfortable with the Committee recognizes that risk adjustment for SDS factors is a rapidly progressing area and that more work is needed to appreciate the effects of social risk, understand the most relevant patient- and community level risk factors, collect data on these risk factors, and determine the best methods to incorporate these risk factors into performance measures.

The Committee stressed the high risk of unintended consequences related to adjustment of these measures for SDS factors and the need to reevaluate these measures as the field continues to move forwards. The Committee recognized the need to ensure facilities serving vulnerable populations are not penalized unfairly while at the same time balancing concerns about worsening healthcare disparities. The Committee looks forward to continued deliberations on these issues and to reexamining these measures as better data emerges. The Committee recommends a reassessment of the availability of SDS variables and a reexamination of these measures through the NQF annual update process.

- While the measure that was submitted to NQF has an Intra-Class Correlation Coefficient below 0.60, the Committee believes it represents an acceptable benchmark for reliability for measurement of excess days in acute care after hospitalization for AMI. The Committee concluded that developers' current approach to risk-adjustment and exclusions met the Scientific Acceptability criteria, and were satisfied with the measure's reliability.
- The Committee followed NQF's guidance on measure harmonization throughout the evaluation process. Prior to the in-person meeting, the Committee received materials regarding these competing measures, and held a separate call after the in-person meeting on September 1 to discuss harmonization issues and allow the developers to answer questions from Committee members. The Committee then voted via survey to recommend both measures. The Committee considered the added value and

2881 Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)
burden of recommending both measures and agreed that the differences in measure specifications added sufficient value to offset any potential negative impact.
7. Consensus Standards Approval Committee (CSAC) Vote: Y-12; N-0 **Decision: Approved for endorsement**
8. Board of Directors Vote: Yes (December 8, 2016) **Decision: Ratified for endorsement**
9. Appeals **Appeal received**

APPENDIX C: Background Information on the Measure Review

0330: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following heart failure (HF) hospitalization

NQF #0330 underwent maintenance evaluation during the All-Cause Admissions and Readmissions Project 2015-2017. The Standing Committee discussed the two updates to the measure. First, the updated measure excluded patients who have either an LVAD or a heart transplant during their indexed stay or during the year prior. The Standing Committee generally agreed that this change was an appropriate reflection of a change in clinical practice. Second, the measure had modest changes to the planned readmissions algorithm which excludes scheduled or planned readmissions from the measure. The Standing Committee noted that there is still a performance gap, with the average heart failure readmission rate over 22 percent and rates ranging from 16 percent to over 32 percent. The Standing Committee was concerned that the published literature suggested a nominal but significant inverse correlation between readmissions and mortality and recommended continued monitoring.

The Committee had extensive conversations about the need to include social risk factors in the risk adjustment model of this measure. The measure developer considered a number of variables related to sociodemographic status (SDS) for potential inclusion in the risk-adjustment model. Candidate SDS variables were selected for examination based on a review of literature and national data sources. SDS variables were ultimately not included in model as the developer found that the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.

A number of public comments were received expressing concerns that SDS variables were not included in the risk adjustment model. The Committee took these concerns seriously and recognizes that there continue to be limitations in the available data elements to capture unmeasured clinical and social risk. Ultimately, the Standing Committee agreed that the measure continues to meet the NQF criteria and recommended NQF #0330 for endorsement.

0506: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following pneumonia hospitalization

NQF #0506 underwent maintenance evaluation during the All-Cause Admissions and Readmissions Project 2015-2017. The Standing Committee reviewed the two measure updates. First, the measure had an expanded cohort including patients who have a principal diagnosis of sepsis and a secondary diagnosis of pneumonia that is present on admission, and patients who have a principal diagnosis of aspiration pneumonia. Second, the measure included the updated planned readmissions algorithm noted above.

The Committee expressed similar concerns about the potential need for social risk factors in the risk adjustment model of this measure. Again, the developer chose not to include social risk factors in the model as the the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.

Public comments also expressed concern about the lack of of adjustment for social risk factors in this measure. Again, the Committee stressed the limitations to currently available data to measure social risk and the need to continue to monitor this issue as the field progresses.

1789 Hospital-Wide All-Cause Unplanned Readmission Measure (HWR)

NQF #1789 underwent maintenance evaluation during the All-Cause Admissions and Readmissions Project 2015-2017. The Standing Committee agreed that there continues to be a performance gap, with all-cause readmission rates ranging from 11.4 percent to 20.1 percent and an average rate of 15.4 percent. The Standing Committee raised concerns that merging multiple cohorts into one group may mask the individual variance properties of the individual cohorts. As with the measures above, the Committee had extensive deliberations on the need to include social risk factors in the risk adjustment model. Again, the developer chose not to include social risk factors in the model as the the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.

Public comments also expressed concern about the lack of of adjustment for social risk factors in this measure. Again, the Committee stressed the limitations to currently available data to measure social risk and the need to continue to monitor this issue as the field progresses.

1891: Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization

NQF #1891 underwent maintenance evaluation the All-Cause Admissions and Readmissions Project 2015-2017. The Standing Committee agreed that the measure continues to have a performance gap with readmission rates for COPD ranging from 15.5 percent to 26.6 percent and an average rate of 20.2 percent. While there was discussion about the modest results of the reliability testing and the use of hierachical logistical modeling, the Standing Committee agreed that the measure met the criteria for NQF endorsement.

As with the measures above, the Committee had extensive deliberations on the need to include social risk factors in the risk adjustment model. Again, the developer chose not to include social

risk factors in the model as the the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.

Public comments also expressed concern about the lack of of adjustment for social risk factors in this measure. Again, the Committee stressed the limitations to currently available data to measure social risk and the need to continue to monitor this issue as the field progresses.

2881: Excess days in acute care (EDAC) after hospitalization for acute myocardial infarction (AMI)

NQF #2881 is a newly-submitted measure that aims to provide a more complete understanding of the quality of care transitions for patients with AMI by measuring a return to acute care after hospital discharge through a number of outcomes: emergency department (ED) visits, observation stays, and unplanned readmissions. This measure is not currently publicly reported, but was finalized for use in CMS' Hospital Inpatient Quality Report (IQR) program starting in FY 2018. The Standing Committee recommended the measure for endorsement, with moderate certainty that the measure scores are reliable and valid, with an intraclass correlation coefficient of 0.54, and a correlation with readmissions of 0.61.

As with the measures above, the Committee had extensive deliberations on the need to include social risk factors in the risk adjustment model. Again, the developer chose not to include social risk factors in the model as the the effect size of each of these variables is small, the c-statistic (i.e., predictive value) is unchanged with the addition of any of these variables into the model, and the addition of any of these variables into the model has little to no effect on hospital performance.

Public comments also expressed concern about the lack of of adjustment for social risk factors in this measure. Again, the Committee stressed the limitations to currently available data to measure social risk and the need to continue to monitor this issue as the field progresses.